

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 04:20:30 ; Search time 1536 seconds
(without alignments)
8545.160 Million cell updates/sec

Title: US-09-622-978-2
Perfect score: 451
Sequence: 1 agaacagaataattattg.....acacacttggcaggttca 451

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.man.*
- 37: em.htg.vit.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	451	100.0	451	6	AX018797	AX018797 Sequence
2	439	97.3	7387	6	AX018796	AX018796 Sequence
3	428	94.9	7389	14	BS7002234	AJ002234 Banana st
4	426.4	94.5	1292	14	AF215816	AF215816 Banana st
5	426.4	94.5	15494	8	AF106946	AF106946 Musa x pa
6	110.8	24.6	1287	14	AF215815	AF215815 Banana st
7	110.8	24.6	1322	6	AX031232	AX031232 Sequence
C 8	58	12.9	15494	8	AF106946	AF106946 Musa x pa
C 9	38.8	8.6	55095	8	AC073506	AC073506 Arabidops
10	38.8	8.6	168259	2	AC067842	AC067842 Homo sapi
11	38.8	8.6	171861	2	AC104824	AC104824 Homo sapi
12	38.8	8.6	192886	9	AC093887	AC093887 Homo sapi
C 13	38.8	8.6	197802	2	AC106879	AC106879 Homo sapi
C 14	37	8.2	72809	2	AC080111	AC080111 Homo sapi
C 15	36.8	8.2	184132	2	AC120345	AC120345 Mus muscu
C 16	36.8	8.2	184672	2	AP002508	AP002508 Homo sapi
17	36.8	8.2	192104	2	AC090229	AC090229 Homo sapi
18	36.8	8.2	194076	2	AC131241	AC131241 Mus muscu
C 19	36.8	8.2	197069	2	AC090370	AC090370 Homo sapi
20	36.8	8.2	207058	2	AP001394	AP001394 Homo sapi
21	36.4	8.1	156547	2	AC112611	AC112611 Rattus no
C 22	36.2	8.0	99512	9	AC004606	AC004606 Homo sapi
C 23	36.2	8.0	160310	9	AC010157	AC010157 Homo sapi
C 24	36	8.0	156593	9	AC078999	AC078999 Homo sapi
C 25	35.8	7.9	205209	2	AC121066	AC121066 Oryctolag
C 26	35.6	7.9	109524	9	HS305316	AL049792 Human DNA
C 27	35.6	7.9	116650	2	AC095948	AC095948 Rattus no
C 28	35.6	7.9	123201	2	AP004615	AP004615 Oryza sat
C 29	35.6	7.9	166915	2	AC102478	AC102478 Mus muscu
C 30	35.6	7.9	171700	2	AC094677	AC094677 Rattus no
31	35.4	7.8	147480	9	AC091639	AC091639 Homo sapi
C 32	35.4	7.8	152843	9	AC025828	AC025828 Homo sapi
33	35.4	7.8	171001	2	AC023894	AC023894 Homo sapi
34	35.4	7.8	179604	9	AC112128	AC112128 Homo sapi
35	35.4	7.8	180333	2	AC021695	AC021695 Homo sapi
36	35.4	7.8	205770	2	AC102597	AC102597 Mus muscu
37	35.2	7.8	124829	9	AC087109	AC087109 Homo sapi
C 38	35	7.8	145905	9	AC087711	AC087711 Homo sapi
C 39	35	7.8	148228	2	AC051659	AC051659 Homo sapi
C 40	35	7.8	167320	2	AF298853	AF298853 Homo sapi
41	35	7.8	170340	2	AC092703	AC092703 Homo sapi
42	34.8	7.7	110000	2	AC099209_2	Continuation (3 of
C 43	34.8	7.7	187209	9	AC090578	AC090578 Homo sapi
C 44	34.8	7.7	187967	2	AC026089	AC026089 Homo sapi
C 45	34.8	7.7	198583	2	AL845423	AL845423 Mus muscu

ALIGNMENTS

RESULT 1
AX018797
LOCUS AX018797 451 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 2 from Patent WO943836.
ACCESSION AX018797
VERSION AX018797.1 GI:10042919
KEYWORDS banana streak virus.
SOURCE banana streak virus.
ORGANISM banana streak virus
REFERENCE 1 (bases 1 to 451)
AUTHORS Viruses; Retroid viruses; Caulimoviridae; Badnavirus.
TITLE Harper, G. and Hull, R.
JOURNAL Banana streak virus promoter and detection
Patent: WO 943836-A 2 02-SEP-1999;
HARPER GLYN (GB); HULL ROGER (GB); PLANT BIOSCIENCE LIMITED (GB)

[illegible][illegible]

```
/codon_start=1
/product="small protein of unknown function"
/protein_id="CAA05263.1"
/db_xref="GI:3183636"
/db_xref="SPTREMBL:O72699"
/translation="MSNSITSSVYQAIAGTIGDWSPGVGSDRGSVNNVTOLITROL
NTIPLCTQEQVIAKLDIVADYONERLIRERTGATSGATPQLKGEIDAINKLSRI
QOIQSQPRKDGGAATSKVQDPYKLRNLK"
1451..6949
/gene="ORF3"
/label="6949"
/label="ORF3"
/codon_start=1
/product="polyprotein, cleavage products include viral
coat protein and proteins with homology to an aspartic
protease, reverse transcriptase and RNase H"
/protein_id="CAA05264.1"
/db_xref="GI:3183637"
/db_xref="SPTREMBL:O72700"
/translation="MAQPRITGSRITTAERGTLIDDOIREYSSRAAYEAQRIAR
OTGNTIGRVGRQPREHTLSMVDPNSPESLSLAHRTVPGVLYMTQDSDPNRIY
RNRTERMLVNGQDQRFIYQESFEELASAGFYIHGLVLOVRIOIMHYIYAGTAL
IVFRTRWTQEEGRSIIIGAMEADLSQHOLIYIPIIDIMTIIRDYQHVQISILTKG
YQFQGEANLLITRSCRLTNPNVGFAYNIQWVLYNSKGVKAQOKLSITKYQ
GTENLKPSNVVPMQPTNLITRVNDNSIRIRGNTQASTSSAPPVNEEDGSDDDI
QATIEHVMYLIEDSDTDYPMVAEEEIFPLENMVEDDIIISOFLNLDITODEDS
RSQVNNLEDEEFFQLREIEKVLSSVAETAISYRPPDAEMGEAPAYAPASGAW
AGSRFFPMQSGPRRWDNSNFIYSLPPAQSQGAMFVPMDFDIKVFWRWESITLH
MTEKIFDNADCKRYMNTLDEDEKHFIEWRMKYTABETMKAQALGDQGTQNIHQ
IRLIFLEPNQVGTTSQDAAYKILSLVCTEMDTATIRYNDYFHSKATNGRAMS
BELSDEFTPLKRGDEVERAFMEKHPSTVGTARTITFKYKLSICRKLVAQKSI
GMDCRSPTRVGLYDYSRYKYGARKSTYKPKSHVRKSHVIRKLSIKRNCRCY
ACGEGRHFASECKNPKIMDRVKVLDLSDLEDGLVLSVGFEDDDVDIYSIDREAN
YRFTNEMEGFNKYIYVIMMEEDPREYLVGPSEWRSMKVSRCQYFCKHKEE
ETHVTICKAGEAPAKHRIDCKEMTQVLCMQPWFYKNVNTVEEVAKIRBKEID
WEDIALKQHEVLTIANEKQLESEVEILIKRQKELKEPIIFEEDETEFAQIKL
EDVERNELNLIQKKEDEIYQNEIIEIKRLEKLEQKQKDEQVNYLEVSIN
ALRPNRNLNICEKEVKNKVVNLNADITGATCVADERMIPSGMEKQAKNKIIRG
VNGTVBNVTAGLWGVKQWYLPQTFIMPSLADGVHMIIGMNFRTVGLRTENGE
VRIYKMTVQAPPIVHELVYIDLELELHEVYNICAESSRGEISEFSPDILIGM
KVLGPSSHRITAMVESGEVDPMTQKRRGERLVNTKRLNDNTKQOYSLPGI
NTIIRKGNAKIYKDFKSGHFQVAMPESIPWFAWIDGLEWYLVMPGLKANAPA
ITQRKMDNFRGTEFIAYYIDILVSETIHOHKEHLKFTCEKNGLVSTPKK
IGTRDIFLGATIGNSIKLPHIHKIIEKMEDEBELKEVGLKWLGLINLARIYIPK
LGLILPLAKTSPNERMNTQDMKIVKEVAVANPELELPEKAIMIYEDGCM
EGGVCKWKTSLQPSWSEKICAVASGKFTPIKSTIDAEIQAVINSLDKFIYLDK
KELIIRIDQALVSEFKKSSDHKPSRWLAFTDYITGTGLEIKFEHIDGKNDVLDI
LSRLVLIILHPKHQSEGVLINAVVEVFKGNTDAQRVNDVVKRYEDWLSKGLHLQ
INVLITSEEPVFCGKNPKAKLKISRTSRNPREFYSCETNCTFTWVWKKQIDSFVQE
KIRWEKLEISEDSLWESLREQENLRAEQBYLIEDALDLIDLSND"
BASE COUNT 2618 a 1357 c 1681 g 1733 t
ORIGIN
Query Match 94.9%; Score 428; DB 14; Length 7389;
Best Local Similarity 99.6%; Pred. No. 8.5e-121; Indels 2; Gaps 2;
Matches 450; Conservative 0; Mismatches 0;
1 AGAACAGATATCTTATTGAAGATGCTCTAGATCTGCTGGATATCAGTAATGATGACTG 60
6889 AGAACAGATATCTTATTGAAGATGCTCTAGATCTGCTGGATATCAGTAATGATGACTG 6948
QY 61 AAGCGGAAGTGGCGGACCCCTACACGCTTGTGATACCAACCGGTGCGAAGCTGATAA 120
6949 AAGCGGAAGTGGCGGACCCCTACACGCTTGTGATACCAACCGGTGCGAAGCTGATAA 7008
DB 121 TGCGGAGTGAGCTGGATACCACTCCTTTATGTAAAGAGAGAGACAAAGTATATGTCTCT 180
7009 TGCGGAGTGAGCTGGATACCACTCCTTTATGTAAAGAGAGAGACAAAGTATATGTCTCT 7068
QY 181 TTTATTTTAAGTTTTCGGTGT-CGTTGCTAGTCACGACGATGACCTTTAGTGAACCTTT 239
7069 TTTATTTTAAGTTTTCGGTGTGCTGTAGTCACGACGATGACCTTTAGT-AACTTT 7127
```

```
QY 240 GCAGGATCTTACGCAAAAGTTGTTAGCCAGAGACATGTCATGATGCTTATCTGCATTAT 299
DB 7128 GCAGGATCTTACGCAAAAGTTGTTAGCCAGAGACATGTCATGATGCTTATCTGCATTAT 7187
QY 300 TGGTGGATGCCACCTTAACGATGCCAGAAAGCTCCACAACTCTCTATATAAGGAGCCTTGT 359
DB 7188 TGGTGGATGCCACCTTAACGATGCCAGAAAGCTCCACAACTCTCTATATAAGGAGCCTTGT 7247
QY 360 ATTACAGGTTGCAACACACGACACCAACCGAGTTTACTCTCTGATTGAGAAATAAAAACT 419
DB 7248 ATTACAGGTTGCAACACACGACACCAACCGAGTTTACTCTCTGATTGAGAAATAAAAACT 7307
QY 420 TCTGTGCTTGAACACACACITTTGTGCGAGTTCA 451
DB 7308 TCTGTGCTTGAACACACACITTTGTGCGAGTTCA 7339
RESULT 4
AF215816 1292 bp DNA linear VRL 03-AUG-2000
DEFINITION Banana streak virus ORF III polyprotein gene, partial cds.
ACCESSION AF215816
VERSION AF215816.1 GI:9255780
KEYWORDS banana streak virus.
SOURCE banana streak virus.
ORGANISM
VIRUSES; Retroid viruses; Caulimoviridae; Badnavirus.
REFERENCE
1 (bases 1 to 1292)
Geering A.D., McMichael, L.A., Dietzgen, R.G. and Thomas, J.E.
Genetic diversity among Banana streak virus isolates from Australia
Phytopathology 90, 921-927 (2000)
JOURNAL
2 (bases 1 to 1292)
Geering A.D. and Thomas, J.E.
Direct Submission
AUTHORS
Submitted (13-DEC-1999) Queensland Horticulture Institute,
Department of Primary Industries, 80 Meiers Road, Indooroopilly,
Queensland 4068, Australia
LOCATION/Qualifiers
1..1292
/organism="Banana streak virus"
/db_xref="taxon:69577"
/country="Australia; North Queensland"
/notes="from Musa acuminata cv. Red Dacca"
<1..852
/notes="ribonuclease H"
/codon_start=1
/product="ORF III polyprotein"
/protein_id="AAF8631.1"
/db_xref="GI:9255781"
/translation="GCMESGGVCKWVDSIQPRSEKICAYASGKFTPIKSTIDAEI
QAVINSLDKFIYLDKELIIRTDQAVISFYKSSDHKPSRWLAFTDYITGTGL
EIKFEHIDGKNDVLDIIRLVKIIILHPKHQSEGVLINAVVEVFKGNTDAQRVND
VVKRYEDWLSKGYRLHQINVLITSEEPVFCGKNPKAKLKISRTSRNPREFYSCETN
CTFTWVWKKQIDSFVQEKIRWEKLEISEDSLWESLREQENLRAEQBYLIEDALDI
LDISND"
BASE COUNT 421 a 235 c 314 g 322 t
ORIGIN
Query Match 94.5%; Score 426.4; DB 14; Length 1292;
Best Local Similarity 99.3%; Pred. No. 2.5e-120;
Matches 449; Conservative 0; Mismatches 1; Indels 2; Gaps 2;
1 AGAACAGATATCTTATTGAAGATGCTCTAGATCTGCTGGATATCAGTAATGATGACTG 60
792 AGAACAGATATCTTATTGAAGATGCTCTAGATCTGCTGGATATCAGTAATGATGACTG 851
QY 61 AAGCGGAAGTGGCGGACCCCTACACGCTTGTGATACCAACCGGTGCGAAGCTGATAA 120
852 AAGCGGAAGTGGCGGACCCCTACACGCTTGTGATACCAACCGGTGCGAAGCTGATAA 911
DB 121 TGCGGAGTGAGCTGGATACCACTCCTTTATGTAAAGAGAGAGACAAAGTATATGTCTCT 180
181 TTTATTTTAAGTTTTCGGTGT-CGTTGCTAGTCACGACGATGACCTTTAGTGAACCTTT 239
7069 TTTATTTTAAGTTTTCGGTGTGCTGTAGTCACGACGATGACCTTTAGT-AACTTT 7127
```

```

Db 912 TCGCGAGTGCAGTGATACCACTACCTTTATGTAAAGAGGAGACAAAGTATAATGTCTCT 971
QY 181 TTATTTTAAAGTTTTCGCGTGT-CGTTGCTAGTCACGACGATGACCTTTAGTGAACCTTT 239
Db 972 TTATTTTAAAGTTTTCGCGTGTGTCTAGTCACGACGATGACCTTTAGTGAACCTTT 15257
QY 240 GCAGGATCTTACGCAAAAGTTTGTAGCCAGAGACATGTGATGCTTATCTGCAATAT 299
Db 1031 GCAGGATCTTACGCAAAAGTTTGTAGCCAGAGACATGTGATGCTTATCTGCAATAT 15317
QY 300 TGGTGGATGCCACCTAACGATGCCAGAAAGCTCCACAACTCTCTATATAAGGAGCCTTGT 359
Db 1091 TGGTGGATGCCACCTAACGATGCCAGAAAGCTCCACAACTCTCTATATAAGGAGCCTTGT 15317
QY 360 ATTCAGGTTGCAAAACGACGACCAACGCGAGTTTACTCTGATTTGAGAAATAAAACT 419
Db 1151 ATTCAGGTTGCAAAACGACGACCAACGCGAGTTTACTCTGATTTGAGAAATAAAACT 15437
QY 420 TCTGTGCTTGAACACACTTTGTGCGAGTTCA 451
Db 1211 TCTGTGCTTGAACACACTTTGTGCGAGTTCA 1242

RESULT 5
AF106946
LOCUS Musa x paradisiaca 15494 bp DNA linear PLN 15-MAR-1999
DEFINITION Musa x paradisiaca clone Musaf banana streak virus sequence.
ACCESSION AF106946
VERSION AF106946.1 GI:4416331
KEYWORDS
SOURCE
ORGANISM Musa x paradisiaca.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Zingiberales; Musaceae;
Musa.
REFERENCE 1 (bases 1 to 15494)
AUTHORS Ndowora,T., Dhanal,G., LaFleur,D., Harper,G., Hull,R., Olszewski,N.
and Lockhart,B.
TITLE Evidence that badnavirus infection in Musa can originate from
integrated pararetroviral sequences
JOURNAL Virology (1999) In press
REFERENCE 2 (bases 1 to 15494)
AUTHORS Olszewski,N.E.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-1998) Plant Biology, University of Minnesota,
1445 Gortner Ave, St. Paul, MN 55108-1095, USA
FEATURES
source
1. .15494
/organism="Musa x paradisiaca"
/cultivar="Obino L'Ewai"
/db_xref="taxon:89151"
/clone="Musaf"
/repeat_region 1613..15494
BASE COUNT 4736 a 3085 c 3110 g 4563 t
ORIGIN
Query Match 94.5%; Score 426.4; DB 8; Length 15494;
Best Local Similarity 99.3%; Pred. No. 2.7e-120;
Matches 449; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 1 AGAACAGATATCTTATTGAAGTCTCTAGATCTGCTGGATATCAGTATGATGACTG 60
Db 15019 AGNACAGATATCTTATTGAAGTCTCTAGATCTGCTGGATATCAGTATGATGACTG 15078
QY 61 AGCGGAATGGCGGACCCCTACCAGGTGTGTATACCAACCGGTGTGAAGCTGATAGA 120
Db 15079 AGCGGAATGGCGGACCCCTACCAGGTGTGTATACCAACCGGTGTGAAGCTGATAGA 15138
QY 121 TCGGAGTGCAGTGGATACCTACTTATGTAAAGAGGAGACAAAGTATAATGTCTCT 180
Db 15139 TCGGAGTGCAGTGGATACCTACTTATGTAAAGAGGAGACAAAGTATAATGTCTCT 15198

```

```

QY 181 TTATTTTAAAGTTTTCGCGTGT-CGTTGCTAGTCACGACGATGACCTTTAGTGAACCTTT 239
Db 15199 TTATTTTAAAGTTTTCGCGTGTGTCTAGTCACGACGATGACCTTTAGTGAACCTTT 15257
QY 240 GCAGGATCTTACGCAAAAGTTTGTAGCCAGAGACATGTGATGCTTATCTGCAATAT 299
Db 15258 GCAGGATCTTACGCAAAAGTTTGTAGCCAGAGACATGTGATGCTTATCTGCAATAT 15317
QY 300 TGGTGGATGCCACCTAACGATGCCAGAAAGCTCCACAACTCTCTATATAAGGAGCCTTGT 359
Db 15318 TGGTGGATGCCACCTAACGATGCCAGAAAGCTCCACAACTCTCTATATAAGGAGCCTTGT 15377
QY 360 ATTCAGGTTGCAAAACGACGACCAACGCGAGTTTACTCTGATTTGAGAAATAAAACT 419
Db 15378 ATTCAGGTTGCAAAACGACGACCAACGCGAGTTTACTCTGATTTGAGAAATAAAACT 15437
QY 420 TCTGTGCTTGAACACACTTTGTGCGAGTTCA 451
Db 15438 TCTGTGCTTGAACACACTTTGTGCGAGTTCA 15469

RESULT 6
AF215815
LOCUS Banana streak virus ORF III polyprotein gene, partial cds.
DEFINITION AF215815
ACCESSION AF215815
VERSION AF215815.1 GI:9255778
KEYWORDS banana streak virus.
SOURCE banana streak virus.
ORGANISM Viruses; Retroid viruses; Caulimoviridae; Badnavirus.
REFERENCE 1 (bases 1 to 1287)
AUTHORS Geering,A.D., McMichael,L.A., Dietzgen,R.G. and Thomas,J.E.
TITLE Genetic diversity among Banana streak virus isolates from Australia
JOURNAL Phytopathology 90, 921-927 (2000)
REFERENCE 2 (bases 1 to 1287)
AUTHORS Geering,A.D. and Thomas,J.E.
TITLE Direct Submission
JOURNAL Submitted (13-DEC-1999) Queensland Horticulture Institute,
Department of Primary Industries, 80 Meiers Road, Indooroopilly,
Queensland 4068, Australia
FEATURES
Location/Qualifiers
1. 1287
/organism="banana streak virus"
/db_xref="taxon:59577"
/country="Australia: North Queensland"
/note="from Musa acuminata cv. Cavendish"
<1..831
/note="ribonuclease H"
/codon_start=1
/product="ORF III polyprotein"
/protein_id="AAF66310.1"
/db_xref="GI:9255779"
BASE COUNT 428 a 236 c 327 g 296 t
ORIGIN
Query Match 24.6%; Score 110.8; DB 14; Length 1287;
Best Local Similarity 67.4%; Pred. No. 5.2e-23;
Matches 260; Conservative 0; Mismatches 117; Indels 9; Gaps 7;

QY 60 GAAGCGGAAGTGGCGGACCCCTACCAGTGTGTATACCAACCGGTGTGA-AGACTGATA 117
Db 840 GTAGCGGAAGTGGCGGACCCCTACCAGTGTGTATACCAACCGGTGTGAAGACTGACG 899
QY 118 AGATCGGAGTGCAGTGGATACCTACTTATGTAAAGAGGAGACAAAGTATAATGTCT 177
Db 900 AGATCGGAGTGCAGTGGATACCTACTTATGTAAAGAGTGGTCTGCTGCCACTACT 959

```

```

QY 178 TCTTTATTTAAAGTTTGTGGTGTCTGTCTAGTCAGCGACGATGACCTTTAGTGAAC 237
Db 960 ---CACATATAGTCTGTCTGAGTGTGCTGTCTGTCACGCAAGA-CCTTAGATTCCT 1015
QY 238 TTGCAGATTCCTACGCAAGTTTGTAGCCAGACAGATGTGATGCTTATCTGCAAT 297
Db 1016 TTGCGTGTAGATGATGCAAGACAGTGTGTCCAGAGTGTGCTGTGACGCGTCCCTTGCATT 1075
QY 298 ATTGGTGGATGCCCTTACGATGTCAGAAAGCTCCACAACCTCTCTATATAGGAGCCTT 357
Db 1076 ATTGGTGGTG-CACCTACGATGCGGAAGCGCACTCCCTCTATAAATAGGACCCCGT 1134
QY 358 GTATTCAGTTGCAACAGCCACCAACGCGAGTTTACTCTCTGATTTGAGAAATAAAA 417
Db 1135 GTATTC-A-GTTGCAAGCAGCAACCAACGCGAGTTTACTCTGAGAGAAATAAGAACA 1193
QY 418 CTTCTGTGCTTGAACACACTTTGTG 443
Db 1194 ATT-TGTGCTTGAATACACTTTGTG 1218

RESULT 7
AX031232
LOCUS 1322 bp DNA linear PAT 20-SEP-2000
DEFINITION Sequence 2 from Patent WO9900492.
ACCESSION AX031232
VERSION AX031232.1 GI:10278576
KEYWORDS
SOURCE
ORGANISM
Badnavirus.
Badnavirus
Viruses; Caulimoviridae.
REFERENCE
1 (bases 1 to 1322)
Dietzen, R.G., Elliott, A.R., Grof, C.P., Thomas, J.E., Geering, A.D.,
McMichael, L.A., Schenk, P.M., Swennen, R.L., Sagi, L. and Remy, S.
Plant and viral promoters
Patent: WO 9900492-A 2 07-JAN-1999;
DIETZEN RALF GEORG (AU) ; ELLIOTT ADRIAN ROSS (AU) ; GROF
CHRISTOPHER PETER LESLIE (AU) ; THOMAS JOHN EDWIN (AU) ; UNIV
QUEENSLAND (AU) ; COMMON SCIENT IND RES ORG (AU) ; GEERING ANDREW
DAVID WILLIAM (AU) ; MCMICHAEL LEE ANNE (AU) ; OF SUGAR EXPERIMENT
STATIONS B (AU) ; SCHENK PEER MARTIN PHILIPP (AU) ; STATE OF
QUEENSLAND ACTING THR (AU) ; SWENNEN RONY LEON (BE) ; SAGI LASZLO
(BE) ; REMY SERGE (BE) ; UNIV LEUVEN KATH (BE)
FEATURES
Location/Qualifiers
1..1322
/organism="Badnavirus"
/isolater="WILLIAMS"
/db_xref="taxon:10652"
BASE COUNT 441 a 239 c 335 g 307 t
ORIGIN

Query Match 24.6%; Score 110.8; DB 6; Length 1322;
Best Local Similarity 67.4%; Pred. No. 5.2e-23;
Matches 260; Conservative 0; Mismatches 117; Indels 9; Gaps 7;

QY 60 GAAGCGAAGTGGCGGACCCC-TACACGCTGTGTATACCAACCGGTGTGA-AGACTGATA 117
Db 858 GTAGCGGAAGTGTATGACCCCACTACCACTGGATGGCACTACCACTGTGACAAGGATACG 917
QY 118 AGATGGGAGTGTGATGATACCACTCTATGTAAGAGGACACAAAGTATATGTC 177
Db 918 AGATGCCAAGTGTGATGATACCACTCTATGTAAGAGGATGCTGTGCTGCACTCACTC 977
QY 178 TCTTTATTTAAAGTTTGTGGTGTCTGTCTAGTCAGCGACGATGACCTTTAGTGAAC 237
Db 978 ---CACTATAGTCTGTCTGAGTGTGCTGTCTGTCACGCAAGA-CCTTAGATTCCT 1033
QY 238 TTGCAGATTCCTACGCAAGTTTGTAGCCAGACAGATGTGATGCTTATCTGCAAT 297
Db 1034 TTGCGTGTAGATGATGCAAGACAGTGTGTCCAGAGTGTGCTGTGACGCGTCCCTTGCATT 1093
QY 298 ATTGGTGGATGCCACCTACGATGTCAGAAAGCTCCACAACCTCTCTATATAGGAGCCTT 357

```

RESULT 8

```

AF106946/c
LOCUS 15494 bp DNA linear PLN 15-MAR-1999
DEFINITION Musa x paradisiaca clone Musaf banana streak virus sequence.
ACCESSION AF106946
VERSION AF106946.1 GI:4416331
KEYWORDS
SOURCE
ORGANISM
Musa x paradisiaca.
Musa x paradisiaca
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Zingiberales; Musaceae;
Musa.
REFERENCE
1 (bases 1 to 15494)
Adowora, T., Dahan, G., LaFleur, D., Harper, G., Hull, R., Olszewski, N.
and Lockhart, B.
Evidence that badnavirus infection in Musa can originate from
integrated pararetroviral sequences
Virology (1999) In press
REFERENCE
2 (bases 1 to 15494)
Olszewski, N.E.
Direct Submission
TITLE
Submitted (17-NOV-1998) Plant Biology, University of Minnesota,
1445 Gortner Ave, St. Paul, MN 55108-1095, USA
JOURNAL
Location/Qualifiers
1..15494
/organism="Musa x paradisiaca"
/cultivar="Obino L'Ewai"
/db_xref="taxon:89151"
/clone="Musaf"
repeat_region 1613..15494
BASE COUNT 4736 a 3085 c 3110 g 4563 t
ORIGIN

Query Match 12.9%; Score 58; DB 8; Length 15494;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 394 TACTCTGATTGAGAAATAAAACTTCTGTGCTTGAACACACTTTGTGCGAGTTCA 451
Db 8139 TACTCTGATTGAGAAATAAAACTTCTGTGCTTGAACACACTTTGTGCGAGTTCA 8082

RESULT 9
AC073506/c
LOCUS 55095 bp DNA linear PLN 19-JAN-2001
DEFINITION Arabidopsis thaliana chromosome 1 BAC F12P21 genomic sequence,
complete sequence.
ACCESSION AC073506
VERSION AC073506.11 GI:12320848
KEYWORDS HTG.
SOURCE
ORGANISM
Arabidopsis thaliana.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE
1 (bases 1 to 55095)
Lin, X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.J.,
Ward, D., Maiti, R., Ronning, C.M., Koo, H., Fujii, C.Y., Utterback, T.R.,
Barstead, M.E., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.
Arabidopsis thaliana chromosome 1 BAC F12P21 genomic sequence

```

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Unpublished
2 (bases 1 to 55095)
Town, C.D. and Kaul, S.
Direct Submission
Submitted (20-JUN-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
3 (bases 1 to 55095)
Town, C.D. and Kaul, S.
Direct Submission
Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
On Jan 19, 2001 this sequence version replaced gi:12280752.
Address all correspondence to: atetigr.org

BAC clone F12P21 is from Arabidopsis thaliana chromosome 1
The orientation of the sequence is from SP6 to 17 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge,
<http://CCR-081.mit.edu/GENSCAN.html>), GenemarkHM (Mark Borodovsky,
<http://genemark.biology.gatech.edu/genemark/>), GlimmerA (a variant
of GlimmerM, see Michaela Perle, <http://www.tigr.org/softlab/glimmerm.htm>), and
GeneSplicer (Michaela Perle and Steven Salzberg, contact
mpertea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).
Simple repeats are identified by RepeatMasker (Arian Smit,
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>).
Location/Qualifiers
1. 55095
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F12P21"

FEATURES

source

misc_feature

1
/note=" (TR) 15229 nt beyond this point were not included
in the submitted sequence due to an overlap with another
BAC (T2H7)"
complement(917..938)
/rpt_family="AT-rich"
complement(1501..1551)
/rpt_family="AT-rich"
complement(2821..2874)
/rpt_family="AT-rich"
complement(3293..3359)
/rpt_family="GAAAAA"
complement(3795..3859)
/rpt_family="AT-rich"
4319..5458
/gene="F12P21.1"

gene

/note="similar to hypothetical protein GI:2832643 from
(Arabidopsis thaliana)"

mRNA

<4319..5458

CDS

4319..5458

/codon_start=1

/product="hypothetical protein"

/protein_id="AAG50568.1"

/db_xref="GI:12320858"

/translation="MSYLLRSDPVSRIHPESQSLTSPDHFLLPDSILLILLIPKQVADV
KDLGRCCIVSRFFSLPVFVNLVYDVCVLSDDSSSDENRFRSLNFIASISDAGGA
CGSFSALFRLVFAIFPFFQMLQILGPKKSSSSFDPSAINDFIVTHSFTQVLK

gene

mRNA

CDS

/note="contains similarity to cyclA protein GI:6358548
from (Antirrhinum graniticum)"

complement(<7926..>8900)

/gene="F12P21.1"

complement(7926..8900)

/gene="F12P21.1"

/codon_start=1

/product="hypothetical protein"

/protein_id="AAG50562.1"

/db_xref="GI:12320852"

/translation="MEVDEIELOKHQEOQSRKLOFSESDNTGLMRNNPNPSSRIIRV
SRAGKDRHSKVLTSKLRDEIRLSVATQFVLDRLGFDQSPKSAVEHLNAAAS
DSITDPLNTNFDHLDQONQKKSACSSGTSESSLISRTKREKARERAKTAK
DRKDLONAHSSFTQLLTGGFDQPSNRNWTGSCDFNPVQIIPNSSQEPNHPFS
FVDFYNGISSSSAANGYSSRQSLFNNNNITQORSISSSSSSSPMD
SQSISFMATPPPLDHNHQLPETFDRLYLYGEGNRSSDDKAKERR"

complement(12129..14611)

/gene="F12P21.2"

/note="similar to putative sugar transporter GI:4263781
from (Arabidopsis thaliana)"

complement(join(<12129..12614,12685..13287,13360..13467,
13554..13589,14313..14383,14473..>14611))

/gene="F12P21.2"

complement(join(12129..12614,12685..13287,13360..13467,
13554..13589,14313..14383,14473..>14611))

/gene="F12P21.2"

/codon_start=1

/product="hypothetical protein"

/protein_id="AAG50560.1"

/db_xref="GI:12320850"

/translation="MEGIHGGADESAFKECFSTWKNPYYLRLAFSAGIGLFGY
DTGVISGALLYIRDDKFSVDNRNTWLOEMIVSAVAGAIIVGAAGWANDKLRGSAIL
MADEFLLIGAIIMAAAPNPSLLVGRVFLGVGMASMTAPLYISEASPATIRGALVS
TNGFLTGQFSLYLNIADTDTGTRWMLGIAGIPALLOVLMFTLPESPRWLYRK
GREEKAILRRIYSADVEDQERALKDSVEVEILEEGSSKINKLKCAKTVRGL
IAGVLOVFOFQVGINVWYPTIVQLAGFASNTALLLSLVLAGNRFGLSIYIF
IDRIGRKLIIISLFGVILSIGITGVFTYERHAPAISSLETQRFNNISCDPYKSA
NTNWDCTMLKASSPSCGYCSSPIGKEHFGACWISDVKDLCNENKRLWTRGCS
NFGFALLGLYIIFSPGMGTVPVINSYIPLFRGICGIIATNANLNLIVAQ
SFLSTAIQTSTWTFIFGVISVIALFLVVCVPETKGMPEIEIKMLERRSMERKFW
KKSKSLYKQNSA"

14385..14448

/rpt_family="CATAn"

complement(15977..16036)

/rpt_family="AT-rich"

complement(16067..16098)

/rpt_family="(TA)n"

complement(17322..17359)

/rpt_family="AT-rich"

18100..19621

/gene="F12P21.12"

/note="similar to elongation factor 1 beta GI:1841870 from
(Pimpinella brachycarpa)"

join(18100..18144,18432..18540,18684..18800,18915..19008,
19097..19220,19311..>19621)

/gene="F12P21.12"

join(18458..18540,18684..18800,18915..19008,19097..19220,
19311..19621)

/gene="F12P21.12"

/codon_start=1

/product="elongation factor 1-beta, putative"

/protein_id="AAG50564.1"

/db_xref="GI:12320854"

/translation="MAAFPNI NSDAGLKLDLHLLRSYITGVQASKODITVPAALAK
PPRSQYVNASRWYNIHDLRLISGVAGSGYIVGSAITAEVATPPAASDKDAA
DEEDDDVDFGEGTEERKAAERASVAKSKESKSLVIDIPWDEETDMKK
LEEAVKSIQMEGLFWGASKLVPVGYGKKLQILCTIVDDLVISIDTMEEQULVPEINE

repeat_region

repeat_region

repeat_region

repeat_region

gene

18100..19621

/gene="F12P21.12"

/note="similar to elongation factor 1 beta GI:1841870 from
(Pimpinella brachycarpa)"

join(18100..18144,18432..18540,18684..18800,18915..19008,
19097..19220,19311..>19621)

/gene="F12P21.12"

join(18458..18540,18684..18800,18915..19008,19097..19220,
19311..19621)

/gene="F12P21.12"

/codon_start=1

/product="elongation factor 1-beta, putative"

/protein_id="AAG50564.1"

/db_xref="GI:12320854"

/translation="MAAFPNI NSDAGLKLDLHLLRSYITGVQASKODITVPAALAK
PPRSQYVNASRWYNIHDLRLISGVAGSGYIVGSAITAEVATPPAASDKDAA
DEEDDDVDFGEGTEERKAAERASVAKSKESKSLVIDIPWDEETDMKK
LEEAVKSIQMEGLFWGASKLVPVGYGKKLQILCTIVDDLVISIDTMEEQULVPEINE

YVSCDIVAFNKKCFHEPIRNSF"
 complement(20413..25010)
 /gene="F12P21.4"
 /note="Predicted by genemark hmh"
 complement(join(20413..21068,21157..21400,21577..21845,
 21920..22047,22327..22664,22803..23009,23107..23199,
 23282..23341,23526..23597,23931..24036,24374..24436,
 24528..24644,24744..25010))
 /gene="F12P21.4"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAG50563.1"
 /db_xref="GI:12320853"
 /translation="MASFREFDMCDLRLKPLRLNLSEYVPEKQPLTNFLSLSKV
 VSTSHKLSESPASIDQHLAKSKAVDDWAKLSALISDDPKMSWGICILIG
 TCQESDREFFSYVFNLSHLKNPASSRIVRVASCTISDLTLRLKSFNTKRD
 AVSHAKLILPILKLEDESSEALLEGIVLLSTVILFPFAAFHSNIDKIEAIAISKI
 FSATSNMLKRAHLLALPKAGDEGWSLMMQKLLISINNTVCTAETVYKGTAKIQ
 RLTPGSDPLPGGONGGLDDANSEOLIVRSYALMFCSTNLTYSKTHQVHKH
 ALAVSKYKWAQQTSPSRDMCYVVIOTSHRLKNSDIRSFNLOINIPVGSLSLV
 EVRLVNGSPRAMSPFMGIQELNCAELPALHSALELLCATLKSIRSQLPPAAS
 VVRLVSIFKCSLPERLKIITITLLKSMGIMGMLAQEVINASVDLDQSLSE
 AFDVASKNLSGALLQACSKRKHSGVEANSVFELRPHNLRSIPSLKIASLE
 ALFETLTIGALGSWSRVSVDNLTATNACEGRNAETPHLCPKSTDLVFEQ
 LAALFASLSVSPRVFAFAGLELFTGKLAQKRVAGVCAHMLSLSEVYTHPR
 ALPLDGLTSLNFPESNFGSEKHNTPNLKNLVIAHGDGDLGNRWAKADVPNSNA
 IORTDKTAPKDYEVVSGTQGBGLDVKLSLMEASIGKIESGESDDDDPISLQ
 EGFLSSSSSDSDIES"
 complement(25584..25619)
 /rpt_family="AT-rich"
 complement(25958..25990)
 /rpt_family="AT-rich"

Query Match 8.6%; Score 38.8; DB 8; Length 55095;
 Best Local Similarity 45.9%; Pred No. 0.98;
 Matches 133; Conservative 0; Mismatches 157; Indels 0; Gaps 0;
 QY 135 GATACCACTCATTATGTAAGAGGAGACAAAGTATATGTCCTCTTTATTTAGTTTG 194
 DB 26449 GTTTCGGTGACCTGTGTTCTAGGACATCAAGAAAGTCTCTTTTCTGTTT 26390
 QY 195 TCGGTGTCGTGTAGTCAGCAGCATGACCTTAGTGAACCTTCAGATCTTACGC 254
 DB 26389 ACCGAAGTGCAGTGCCTTTTAATAGTTAGTTAGTTAGTTGATGTCATTATAGC 26330
 QY 255 AAAAGTTGTAGCCAGACATGATGATGATGATGATGATGATGATGATGATGATGATG 314
 DB 26329 TTGGTGTGTTGTTTTCAGAGTCAGTCAAGTTTCTCTCTCTCTCTCTCTCTCTCT 26270
 QY 315 AACGATGCCAGAACTCCACACTCTCTATATAAGAGCCCTGTATTGAGTTGCAAC 374
 DB 26269 GAAATAATTAATCTGTGAACCTGTGTTGTTGAAGATGATGTTCTTATGTTTCAAC 26210
 QY 375 ACGCACCACACGCGAGTTTACTCCIGATTGAGATTAAGAAATAAAACTCTGT 424
 DB 26209 CGAAAGAACATATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGT 26160

RESULT 10
 AC067842
 LOCUS
 DEFINITION
 Homo sapiens chromosome 4 clone RP11-772A14 map 4, WORKING DRAFT
 SEQUENCE, 16 unordered pieces.
 AC067842
 AC067842.2 GI:8099884
 HTGS_PHASE1; HTGS_DRAFT.
 Homo sapiens.

ORGANISM REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS

Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 168259)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G., Collins,S.,
 Campoliano,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
 Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
 Melrlm,J., Meneus,J., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 28, 2000 this sequence version replaced gi:7651893.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center Project name: L9386
 Center clone name: 772_A14
 ----- Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 158765 bases at least Q40
 Consensus quality: 163892 bases at least Q30
 Consensus quality: 165798 bases at least Q20
 Insert size: 170000; agarose-fp
 Insert size: 166759; sum-of-contigs
 Quality coverage: 4.3 in Q20 bases; agarose-fp
 Quality coverage: 4.4 in Q20 bases; sum-of-contigs

* NOTE: this is a 'working draft' sequence. It currently
 * consists of 16 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1161: contig of 1161 bp in length
 * 1162 1261: gap of 100 bp
 * 1262 2768: contig of 1507 bp in length
 * 2769 2868: gap of 100 bp
 * 2869 7059: contig of 4191 bp in length
 * 7060 7159: gap of 100 bp
 * 7160 10594: contig of 3435 bp in length
 * 10595 10694: gap of 100 bp
 * 10695 16254: contig of 5560 bp in length


```
Query Match      8.6%; Score 38.8; DB 2; Length 168259;
Best Local Similarity 54.1%; Pred. No. 1;
Matches 79; Conservative 0; Mismatches 67; Indels 0;
```



```

FEATURES          * 85969 171861: contig of 85993 bp in length.
source            1. 171861
                  Location/Qualifiers
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="4"
                  /clone="RP11-772A14"
                  1. 1480
misc_feature      /note="assembly_name:Contig7"
                  1581..21730
misc_feature      /note="assembly_name:Contig20"
                  21831..85868
misc_feature      /note="assembly_name:Contig21"
                  85969..171861
misc_feature      /note="assembly_name:Contig22
                  clone_end:T7
                  vector_side:right"
BASE COUNT      52804 a 33074 c 32245 g 53434 t 304 others
ORIGIN
Query Match      8.6%; Score 38.8; DB 2; Length 171861;
Best Local Similarity 54.1%; Pred. No. 1;
Matches 79; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
QY 3 AACAGAAATACCTATTGGAAGATGCTAGATGCTGCTGGATATCATGTAATGATGACTGAA 62
DB 118933 AAATGGTAGATATTATTGGTATTCTATAGATGCGACCAAGATATTATTCTCCGTAT 118992
QY 63 CGGAAGTGGCGGACCCCTACCAAGCTGTTGATACCAACCGGTGTGAAGACTGATAGATG 122
DB 118993 TCAAACTGAGGCAAAATCTACCGTGTGTGATACACACAGATGTAGATGGCCAAAATG 119052
QY 123 CGGAGTGAGCTGGGATACCACTCACTT 148
DB 119053 AAGATTAGTTGATCTCTAATCTATT 119078

RESULT 12
AC093887
DEFINITION      Homo sapiens BAC clone RP11-667D12 from 4, complete sequence.
ACCESSION      AC093887
VERSION        AC093887.3 GI:15982604
KEYWORDS       HTG.
SOURCE         Homo sapiens.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
               Sulston, J.E. and Waterston, R.
               1 (bases 1 to 192886)
               Toward a complete human genome sequence
               Genome Res. 8 (11), 1097-1108 (1998)
               99063792
               PUBMED 9847074
REFERENCE      2 (bases 1 to 192886)
               Isak, A., Kozlowski, A., Doebber, A. and Bielicki, L.
               The sequence of Homo sapiens BAC clone RP11-667D12
               Unpublished (2001)
REFERENCE      3 (bases 1 to 192886)
               Waterston, R.H.
               Direct Submission
               Submitted (10-SEP-2001) Genome Sequencing Center, Washington
               University School of Medicine, 4444 Forest Park Parkway, St. Louis,
               MO 63108, USA
               4 (bases 1 to 192886)
               Waterston, R.H.
               Direct Submission
               Submitted (07-OCT-2001) Genome Sequencing Center, Washington
               University School of Medicine, 4444 Forest Park Parkway, St. Louis,
               MO 63108, USA
               5 (bases 1 to 192886)
               Waterston, R.
               Direct Submission

```

JOURNAL

COMMENT

Submitted (01-MAR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 7, 2001 this sequence version replaced gi:15741648.

----- Genome Center

Center: Washington University Genome Sequencing Center
Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: sapiens@wustl.wustl.edu

----- Summary Statistics

Center project name: H_NH0667D12

Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frenken, E., Tateo, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-772A14. Actual start of this clone is at base position 1 of RP11-667D12; actual end is at base position 192886 of RP11-667D12.

Data from AC067842 was used to finish this clone, AC027193. Single stranded regions exist at 99771 and 157730.

The sequence of AC027193 has been incorporated into AC093887.

FEATURES

source

1. 192886

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="4"

/map="4"

/clone="RP11-667D12"

/clone_lib="RPCI-11"

395..615

/rpt_family="L2"

/rpt_family="MaLR"

973..1126

/rpt_family="L2"

1127..1462

/rpt_family="MaLR"

1463..1932

/rpt_family="MaLR"

1933..2101

/rpt_family="MaLR"

2179..2265

/rpt_family="MaLR"

/rpt_family="L2"

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

```

repeat_region 2270. 2316 /rpt_family="Mariner"
repeat_region 2351. 2799 /rpt_family="ERV1"
repeat_region 2801. 3925 /rpt_family="ERV1"
repeat_region 3923. 3954 /rpt_family="ERV1"
repeat_region 3926. 4233 /rpt_family="(T)n"
repeat_region 4234. 4406 /rpt_family="Alu"
repeat_region 4407. 4482 /rpt_family="ERV1"
repeat_region 4488. 4718 /rpt_family="ERV1"
repeat_region 4731. 5091 /rpt_family="L2"
repeat_region 5272. 5419 /rpt_family="MaLR"
repeat_region 7522. 7552 /rpt_family="L2"
repeat_region 8414. 8439 /rpt_family="(TTTA)n"
repeat_region 11042. 11397 /rpt_family="(T)n"
repeat_region 13916. 14374 /rpt_family="L1"
repeat_region 14375. 14842 /rpt_family="L2"
repeat_region 14877. 14948 /rpt_family="MaLR"
repeat_region 14984. 15292 /rpt_family="L2"
repeat_region 15346. 15672 /rpt_family="Alu"
repeat_region 15677. 15839 /rpt_family="MIR"
repeat_region 15924. 15963 /rpt_family="MIR"
repeat_region 15964. 16257 /rpt_family="MaLR"
repeat_region 16064. 16084 /rpt_family="Alu"
repeat_region 16229. 16257 /rpt_family="AT_rich"
repeat_region 16258. 16629 /rpt_family="(CAAAA)n"
repeat_region 18394. 18816 /rpt_family="MaLR"
repeat_region 18944. 18988 /rpt_family="MaLR"
repeat_region 25631. 25698 /rpt_family="AT_rich"
repeat_region 25695. 25762 /rpt_family="MERL_type"
repeat_region 27613. 27652 /rpt_family="MaLR"
repeat_region 30257. 30935 /rpt_family="MIR"
repeat_region 30523. 30589 /rpt_family="ERV1"
repeat_region 30951. 31045 /rpt_family="AT_rich"
repeat_region 31133. 31388 /rpt_family="MIR"
repeat_region 31504. 31816 /rpt_family="L1"
repeat_region 32055. 32096 /rpt_family="Alu"
repeat_region 32100. 32178 /rpt_family="(TA)n"
repeat_region 32156. 32273 /rpt_family="L1"

```

```

repeat_region 32318. 32408 /rpt_family="(TA)n"
repeat_region 33536. 33645 /rpt_family="(TA)n"
repeat_region 34331. 34418 /rpt_family="L2"
repeat_region 34942. 35063 /rpt_family="MIR"
repeat_region 35108. 35549 /rpt_family="L2"
repeat_region 35588. 35653 /rpt_family="ERV1"
repeat_region 35686. 35786 /rpt_family="L2"
repeat_region 35785. 35890 /rpt_family="L2"
repeat_region 36335. 36370 /rpt_family="MIR"
repeat_region 36352. 38205 /rpt_family="AT_rich"

Query Match 8.6%; Score 38.8; DB 9; Length 192886;
Best Local Similarity 54.1%; Pred. No. 1;
Matches 79; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 3 AACAGAAATCTTATGAGATGCTGTAGATCTGCTGGATATCAGTAATGACTGAA 62
Db 35874 AAATGGTAGATATTTGGTATTTCTTAGATGGAACCAAGATATTAATTTCCGTAT 35933
QY 63 GCCGAAGTGGCGGACCCCTACACAGCTGTGATACCAACGGTGAAGACTGATAGATG 122
Db 35934 TCAAACTGAGGCAATCTACCGTGTGTGATACACAGATGTGTAGATGCCAAAATG 35993
QY 123 CGGAGTGAGCTGGATACCACTCACTT 148
Db 35994 AAGAAATAGTGTATCTCACTTATT 36019

```

RESULT 13 AC106879/c

```

LOCUS AC106879 197802 bp DNA linear HTG 14-AUG-2002
DEFINITION Homo sapiens chromosome 4 clone RP11-674B11, WORKING DRAFT
SEQUENCE, 4 unordered pieces.
ACCESSION AC106879
VERSION AC106879.2 GI:22218383
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 197802)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS Waterston,R.H.
REFERENCE 2 (bases 1 to 197802)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 3 (bases 1 to 197802)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Aug 14, 2002 this sequence version replaced gi:18139544.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu

```



```
* 6607 6706: gap of 100 bp
* 6707 7438: contig of 732 bp in length
* 7439 7538: gap of 100 bp
* 7539 8234: contig of 696 bp in length
* 8235 8334: gap of 100 bp
* 8335 9055: contig of 721 bp in length
* 9056 9155: gap of 100 bp
* 9156 9888: contig of 733 bp in length
* 9889 9988: gap of 100 bp
* 10739 10738: contig of 750 bp in length
* 10839 11579: contig of 741 bp in length
* 11580 12404: contig of 725 bp in length
* 12405 13257: gap of 100 bp
* 13258 13357: contig of 753 bp in length
* 13358 14094: contig of 737 bp in length
* 14095 14194: gap of 100 bp
* 14195 14929: contig of 735 bp in length
* 14930 15029: gap of 100 bp
* 15030 15727: contig of 698 bp in length
* 15728 15827: gap of 100 bp
* 15828 16515: contig of 688 bp in length
* 16516 16615: gap of 100 bp
* 16616 17344: contig of 729 bp in length
* 17345 17444: gap of 100 bp
* 17445 18137: contig of 693 bp in length
* 18138 18237: gap of 100 bp
* 18238 19017: contig of 780 bp in length
* 19018 19117: gap of 100 bp
* 19118 19846: contig of 729 bp in length
* 19847 19946: gap of 100 bp
* 19947 20687: contig of 741 bp in length
* 20688 20787: gap of 100 bp
* 20788 21522: contig of 735 bp in length
* 21523 21622: gap of 100 bp
* 21623 22349: contig of 727 bp in length
* 22350 22449: gap of 100 bp
* 22450 23179: contig of 730 bp in length
* 23180 23279: gap of 100 bp
* 23280 24033: contig of 754 bp in length
* 24034 24133: gap of 100 bp
* 24134 24895: contig of 762 bp in length
* 24896 24995: gap of 100 bp
* 24996 25733: contig of 738 bp in length
* 25734 25833: gap of 100 bp
* 25834 26545: contig of 712 bp in length
* 26546 26645: gap of 100 bp
* 26646 27379: contig of 734 bp in length
* 27380 27479: gap of 100 bp
* 27480 28173: contig of 694 bp in length
* 28174 28273: gap of 100 bp
* 28274 28986: contig of 713 bp in length
* 28987 29086: gap of 100 bp
* 29087 29832: contig of 746 bp in length
* 29833 29932: gap of 100 bp
* 29933 30661: contig of 729 bp in length
* 30662 30761: gap of 100 bp
* 30762 31475: contig of 714 bp in length
* 31476 31575: gap of 100 bp
* 31576 32255: contig of 680 bp in length
* 32256 32355: gap of 100 bp
* 32356 33066: contig of 711 bp in length
* 33067 33166: gap of 100 bp
* 33167 33903: contig of 797 bp in length
* 33904 34063: gap of 100 bp
* 34064 34834: contig of 771 bp in length
* 34835 34934: gap of 100 bp
* 34935 35661: contig of 727 bp in length
* 35662 35761: gap of 100 bp
* 35762 36480: contig of 719 bp in length
* 36481 36580: gap of 100 bp
```

```
* 36581 37308: contig of 728 bp in length
* 37309 37408: gap of 100 bp
* 37409 38126: contig of 718 bp in length
* 38127 38226: gap of 100 bp
* 38227 38980: contig of 754 bp in length
* 38981 39080: gap of 100 bp
* 39081 39833: contig of 753 bp in length
* 39834 39933: gap of 100 bp
* 39934 40669: contig of 736 bp in length
* 40670 40769: gap of 100 bp
* 40770 41496: contig of 727 bp in length
* 41497 41596: gap of 100 bp
* 41597 42335: contig of 739 bp in length
* 42336 42435: gap of 100 bp
* 42436 43170: contig of 735 bp in length
* 43171 43270: gap of 100 bp
* 43271 44021: contig of 751 bp in length
* 44022 44121: gap of 100 bp
* 44122 44855: contig of 734 bp in length
* 44856 44955: gap of 100 bp
* 44956 45675: contig of 720 bp in length
* 45676 45775: gap of 100 bp
* 45776 46485: contig of 710 bp in length
* 46486 46585: gap of 100 bp
* 46586 47339: contig of 754 bp in length
* 47340 47439: gap of 100 bp
* 47440 48133: contig of 694 bp in length
* 48134 48233: gap of 100 bp
* 48234 48927: contig of 694 bp in length
* 48928 49027: gap of 100 bp
* 49028 49742: contig of 715 bp in length
* 49743 49842: gap of 100 bp
* 49843 50569: contig of 727 bp in length
* 50570 50669: gap of 100 bp
* 50670 51374: contig of 705 bp in length
* 51375 51474: gap of 100 bp
* 51475 52233: contig of 759 bp in length
* 52234 52333: gap of 100 bp
* 52334 53034: contig of 701 bp in length
* 53035 53134: gap of 100 bp
* 53135 53866: contig of 732 bp in length
* 53867 53966: gap of 100 bp
* 53967 54687: contig of 721 bp in length
* 54688 54787: gap of 100 bp
* 54788 55518: contig of 731 bp in length
* 55519 55618: gap of 100 bp
* 55619 56354: contig of 736 bp in length
* 56355 56454: gap of 100 bp
* 56455 57210: contig of 756 bp in length
* 57211 57310: gap of 100 bp
* 57311 58011: contig of 701 bp in length
```

Query Match 8.2%; Score 37; DB 2; Length 72809;

Best Local Similarity 52.2%; Pred. No. 3.6;

Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
QY 230 AGTGAACCTTGCAGGATCTTACGAAAGTTGTAGCCAGACAGACATGTGATGCTTA 289
    || || || || || || || || || || || || || || || || || || || ||
Db 35557 AGACCATTCTTGGTAATTTGGAAGCAATTTGTTGACCAGGAACGTTGGAGGATACA 35498
    || || || || || || || || || || || || || || || || || || || ||
QY 290 TCTGCATTATTGGTGGATGCCACCTAACGATCCAGCAAGCTCCACAACTCTCTATATA 349
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 35497 TAGAARAGAGGGTTGGATCCCTTTTAAANAATGATAATATCCCCAAAAGGATAAAT 35438
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 350 GGAGCCTTCTATTCAGGTTGCAACACGCCACCAAC 386
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35437 GGAAGATGTTTTGAATGCGGCAAAACTGCCGAAAC 35401
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 15

AC120345/c

LOCUS

DEFINITION Mus musculus clone RP23-330P24, WORKING DRAFT SEQUENCE, 10 ordered

AC120345 184132 bp DNA linear

HTG 24-AUG-2002

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

pieces.
AC120345
AC120345.3 GI:22475041
HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-330P24
Unpublished

REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS

2 (bases 1 to 184132)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,I., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campolano,A., Chang,J.,
Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 184132)

TITLE
JOURNAL
REFERENCE
AUTHORS

Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 24, 2002 this sequence version replaced gi:20503168.
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
JOURNAL
COMMENT

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L20879
Center clone name: 330_P_24
----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 180974 bases at least Q40
Consensus quality: 182307 bases at least Q30
Consensus quality: 182910 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 183232; sum-of-contigs
Quality coverage: 12.6 in Q20 bases; agarose-fp
Quality coverage: 12.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 22383: contig of 22383 bp in length
* 22384 22483: gap of 100 bp
* 22484 24264: contig of 1781 bp in length
* 24265 24364: gap of 100 bp
* 24365 25784: contig of 1420 bp in length
* 25785 25884: gap of 100 bp
* 25885 27850: contig of 1966 bp in length
* 27851 36941: contig of 8991 bp in length
* 36942 37041: gap of 100 bp
* 37042 47790: contig of 10749 bp in length
* 47791 47890: gap of 100 bp
* 47891 108970: contig of 61080 bp in length
* 108971 109070: gap of 100 bp
* 109071 133941: contig of 24871 bp in length
* 133942 134041: gap of 100 bp
* 134042 168684: contig of 34643 bp in length
* 168685 168784: gap of 100 bp
* 168785 184132: contig of 15348 bp in length.

FEATURES

	Location/Qualifiers	source
1..184132	/organism="Mus musculus"	
	/db_xref="taxon:10090"	
	/clone="RP23-330P24"	
	/clone_lib="RPCI-23 Female Mouse BAC"	
1..22383	/note="assembly_fragment"	
	clone_end:SP6	
	vector_side:left	
22484..24264	/note="assembly_fragment"	
24365..25784	/note="assembly_fragment"	
25885..27850	/note="assembly_fragment"	
27951..36941	/note="assembly_fragment"	
37042..47790	/note="assembly_fragment"	
47891..108970	/note="assembly_fragment"	
109071..133941	/note="assembly_fragment"	
134042..168684	/note="assembly_fragment"	
168785..184132	/note="assembly_fragment"	
	clone_end:T7	
	vector_side:right	

BASE COUNT 53226 a 38749 c 37916 g 53338 t 903 others
ORIGIN
Query Match 8.2%; Score 36.8; DB 2; Length 184132;

```
Best Local Similarity 49.5%; Pred. No. 4.3;
Matches 95; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

Qy 164 CAAAGTAAATGCTCTTTTAAAGTTGTCGGTGTGCTGTAGTCAGCGACGANG 223
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
98957 CCACGCATGCTGTCAAAGTCCTTGGGAACATAGAGGAGCCTCAGTCAAGCGACACACACA 98898

Qy 224 ACCTTTAGTGAACCTTTGCAGGATTCCTAGCCTAAAGTTGTTAGGCCAGAGACATGTGATGA 283
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
98897 GCCAGGTGTGATATTTACATACACTCACCTACATTTAAATTCACATCAAGCAACAG 98838

Qy 284 TGCTTATCTGCATATTTGGTGGATGCCACTACAGATGCCAGAAAGCTCCACAACCTCTCT 343
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
98837 TGCATATTTGTTTATTTGGTGCATATGAATGTTCTGAGCCTTGATCACCACACTACCCACT 98778

Qy 344 ATATAAGGAGCC 355
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
98777 ACATGCTAACCC 98766
```

Search completed: June 24, 2003, 05:35:37
Job time : 1541 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 03:07:45 ; Search time 189 Seconds
(without alignments)
5373.816 Million cell updates/sec

Title: US-09-622-978-2

Perfect score: 451

Sequence: 1 agacaagaatatctattg.....acacattgtgcgagtcca 451

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

1: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	437	96.9	7387	20 AAZ20893	Nigerian isolate o
2	428	94.9	451	20 AAZ20894	Banana Streak Viru
3	110.8	24.6	1322	20 AAX06864	Australian banana
C 4	33.4	7.4	6071	22 AAS45372	Chemically pretrea
C 5	33.4	7.4	6071	24 ABK28209	DNA transcription
6	33	7.3	7516	24 ABL32060	Human immune syste
C 7	32.8	7.3	1045	21 AAC79035	Human secreted pro
C 8	32.8	7.3	2257	23 ABL24210	Drosophila melanog
9	32.2	7.1	654	23 ABV14414	Human prostate exp

10	31.8	7.1	555	24	ABQ53082	Oligonucleotide fo
C 11	31.8	7.1	555	24	ABQ53083	Oligonucleotide fo
C 12	31.8	7.1	1180	19	AAK14043	H. pylori GHPO 141
C 13	31.8	7.1	1234	19	AAV31267	E. coli J96 pathog
C 14	31.6	7.0	214	22	AAS56470	Human CDNA for an
C 15	31.6	7.0	460	22	ABA42809	Human breast cell
C 16	31.6	7.0	460	22	ABA53235	Human foetal liver
C 17	31.6	7.0	460	22	ABK01486	Human brain expres
C 18	31.6	7.0	460	22	AAK26933	Human bone marrow
19	31.6	7.0	1497	17	AAK06750	Resynthesised jojo
20	31.6	7.0	4107	23	AAS72414	DNA encoding novel
21	31.6	7.0	10034	24	ABL33884	Human immune syste
22	31.6	7.0	16720	24	ABL32415	Human immune syste
23	31.6	7.0	16720	24	AAS61119	Human gene regulat
24	31	6.9	772	20	AAK17517	Human gene express
25	31	6.9	3145	23	ABL23378	Drosophila melanog
26	31	6.9	16956	21	AAK91923	Wild type (C57BL/6
C 27	30.8	6.8	297	16	AAK26314	Human gene signatu
28	30.8	6.8	910715	20	AAK20248	Borrelia burgdorfe
29	30.4	6.7	1176	21	AAK65106	Membrane-bound pro
30	30.4	6.7	1176	22	AAS46071	Human DNA encoding
31	30.4	6.7	1176	22	AAK92101	Human PRO1270 CDNA
32	30.4	6.7	1176	22	AAK44252	Human PRO1270 (UNQ
33	30.4	6.7	2009	20	AAK99286	Rupestria stem pit
34	30.4	6.7	24079	22	AAK71153	Human immune/naema
35	30.2	6.7	1643	21	AAK56150	Human Cathepsin S
36	30.2	6.7	1763	24	ABK84111	Human GDP-L-fucose
37	30.2	6.7	1763	24	ABL61836	Colon adenocarcino
38	30.2	6.7	3423	24	ABL60222	Human secreted pro
39	30	6.7	244	21	AAK32605	Human prostate exp
40	30	6.7	394	23	ABV37609	CASB414 polynucleo
41	30	6.7	550	20	AAK28483	Oligonucleotide fo
42	30	6.7	902	24	ABQ49966	Oligonucleotide fo
C 43	30	6.7	902	24	ABQ49967	Oligonucleotide fo
44	30	6.7	1499	23	ABV24927	Human prostate exp
45	30	6.7	1893	24	ABK33542	CDNA encoding huma

ALIGNMENTS

RESULT 1
AAZ20893
ID AAZ20893 standard; DNA; 7387 BP.
XX AAZ20893;
XX AAZ20893;
XX 01-DEC-1999 (first entry)
DT 01-DEC-1999 (first entry)
DE Nigerian isolate of Banana Streak Virus DNA.
XX
KW detection; Banana Streak Virus; promoter; nigerian isolate;
KW badnavirus; ds.
XX
OS Banana Streak Virus.
XX
PN WO9943836-A1.
XX
PD 02-SEP-1999.
XX
PF 26-FEB-1999; 99WO-GB00599.
XX
PR 27-FEB-1998; 98GB-0004293.
XX
XX (PLAN-) PLANT BIOSCIENCE LTD.
XX
XX Hull R, Harper G;
XX
XX WPI; 1999-527629/44.
XX
XX New virus promoter, used for the production of transgenic plants, and
XX for identifying antiviral agents and for the detection of the virus -

PS Disclosure; Page 65-68; 78pp; English.

XX This is the complete 7388 bp nucleotide sequence of the Nigerian isolate
CC of BSV. Banana Streak Virus (BSV) is a member of the badnavirus
CC group, which have non-enveloped bacilliform particles of size 30 times
CC 130-150 nm, containing a circular double-stranded DNA genome of 7.4 to
CC 8.0 kbp. This sequence was used to identify the BSV promoter.
CC The Banana Streak Virus (BSV) promoter (AAZ20894) can be used for
CC producing transgenic plants, particularly banana plants and rice and
CC sugar cane.
CC Expression of heterologous nucleic acids can be used to influence
CC characteristics such as resistance, immunity, tolerance,
CC hypersensitivity to pathogens such as viruses, fungi and bacteria,
CC pests such as nematodes and weevils, agronomic characters such as
CC dwarfism of the plant, yield of seed or other product, fertility or
CC sterility and quality of fruit.
CC The promoter constructs can also be used for screening for a substance
CC able to modulate activity of the promoter.
CC For anti-viral purposes, e.g. for treatment of BSV in banana or other
CC disease, a substance able to down-regulate expression of the promoter
CC may be sought.
CC The polynucleotide promoter sequence can also be used for the detection
CC of BSV.

XX Sequence 7387 BP; 2607 A; 1358 C; 1680 G; 1742 T; 0 other;

Query Match 96.9%; Score 437; DB 20; Length 7387;

Best Local Similarity 99.8%; Pred. No. 3.9e-131;

Matches 448; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 3 AACAGAAATATCTTATGTAAGATGCTCTAGATCTGCTGGATATCATGTAATGACTGAA 62

Db - 6890 AACAGAAATATCTTATGTAAGATGCTCTAGATCTGCTGGATATCATGTAATGACTGAA 6949

QY 63 GCGGAAGTGGCGGACCCCTACCACTGTTGATACCAACCGGTGTGAAGCTGATAAGATG 122

Db 6950 GCGGAAGTGGCGGACCCCTACCACTGTTGATACCAACCGGTGTGAAGCTGATAAGATG 7009

QY 123 CGAGTGTAGTGATACCACTCATCTTATGTAAGAGGAGACAAAGTATATGTCCTTT 182

Db 7010 CGAGTGTAGTGATACCACTCATCTTATGTAAGAGGAGACAAAGTATATGTCCTTT 7069

QY 183 ATTTTAAGTTTGTGCGGTGCTGTGTCTAGTACGACGATGACCTTTAGTGAACCTTGCA 242

Db 7070 ATTTTAAGTTTGTGCGGTGCTGTGTCTAGTACGACGATGACCTTTAGTGAACCTTGCA 7128

QY 243 GGATTTCTAGCGAAAGTTGTTAGGCGAGAGACATGTGATGCTTATCTGCATTATGG 302

Db 7129 GGATTTCTAGCGAAAGTTGTTAGGCGAGAGACATGTGATGCTTATCTGCATTATGG 7188

QY 303 TGGATGCCACCTAACCATGCCAAGAGCTCCCAACTCTCTATATAAGAGCGCTTGATT 362

Db 7189 TGGATGCCACCTAACCATGCCAAGAGCTCCCAACTCTCTATATAAGAGCGCTTGATT 7248

QY 363 CAGGTTGCAAAACAGCACCAACGCGGATTTTACCTCGATTTGAGAATAAAACTTCT 422

Db 7249 CAGGTTGCAAAACAGCACCAACGCGGATTTTACCTCGATTTGAGAATAAAACTTCT 7308

QY 423 GTCTTGAACACACATTTGTGCGAGTTCA 451

Db 7309 GTCTTGAACACACATTTGTGCGAGTTCA 7337

RESULT 2

AAZ20894

ID AAZ20894 standard; DNA; 451 BP.

XX AC AAZ20894;

XX 01-DEC-1999 (first entry)

DT Banana Streak Virus promoter nucleotide sequence.

DE

XX

KW detection; Banana Streak Virus; promoter; nigerian isolate;

KW badnavirus; promoter; ds.

XX Banana Streak Virus.

OS WO9943836-A1.

PN 02-SEP-1999.

XX 26-FEB-1999; 99WO-GB00599.

XX 27-FEB-1998; 98GB-0004293.

XX (PLAN-) PLANT BIOSCIENCE LTD.

XX Hull R, Harper G;

XX WPI; 1999-527629/44.

XX New virus promoter, used for the production of transgenic plants, and

XX for identifying antiviral agents and for the detection of the virus -

XX Claim 1; Page 69; 78pp; English.

XX This is the nucleotide sequence of the Banana Streak Virus promoter.

XX The Banana Streak Virus (BSV) promoter (AAZ20894) can be used for

XX producing transgenic plants, particularly banana plants and rice and

XX sugar cane.

XX Expression of heterologous nucleic acids can be used to influence

XX characteristics such as resistance, immunity, tolerance,

XX hypersensitivity to pathogens such as viruses, fungi and bacteria,

XX pests such as nematodes and weevils, agronomic characters such as

XX dwarfism of the plant, yield of seed or other product, fertility or

XX sterility and quality of fruit.

XX The promoter constructs can also be used for screening for a substance

XX able to modulate activity of the promoter.

XX For anti-viral purposes, e.g. for treatment of BSV in banana or other

XX disease, a substance able to down-regulate expression of the promoter

XX may be sought.

XX The polynucleotide promoter sequence can also be used for the detection

XX of BSV.

XX Sequence 451 BP; 131 A; 89 C; 104 G; 127 T; 0 other;

Query Match 94.9%; Score 428; DB 20; Length 451;

Best Local Similarity 99.6%; Pred. No. 9.6e-129;

Matches 450; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1 AGAACAAAGATATCTTATTGAAGATGCTCTAGATCTGCTGGATATCAGTATGATGACTG 60

Db 1 AGAACAAAGATATCTTATTGAAGATGCTCTAGATCTGCTGGATATCAGTATGATGACTG 60

QY 61 AGCGGAAGTGGCGGACCCCTACCACTGTTGATACCAACCGGTGTGAAGACTGATAAGA 120

Db 61 AGCGGAAGTGGCGGACCCCTACCACTGTTGATACCAACCGGTGTGAAGACTGATAAGA 120

QY 121 TGGCGGATGAGTGGATACCACTCATCTTATTGAAGAGGAGACAAAGTATATGTCCTCT 180

Db 121 TGGCGGATGAGTGGATACCACTCATCTTATTGAAGAGGAGACAAAGTATATGTCCTCT 180

QY 181 TTATTTTAAAGTTTGTGCGGTGTCTAGTACGACGATGACCTTTAGTGAACCTTT 239

Db 181 TTATTTTAAAGTTTGTGCGGTGTCTAGTACGACGATGACCTTTAGTGAACCTTT 239

QY 240 GCAGGATCTTACGCAAAAGTTGTTAGGCGAGACATGCTGATGCTTATCTCATATAT 299

Db 240 GCAGGATCTTACGCAAAAGTTGTTAGGCGAGACATGCTGATGCTTATCTCATATAT 299

QY 300 TGGTGGATGCCACCTAACGATGCCAGAAAGTCTCCAACTCTCTATATAAGGAGCCTTGT 359

Db 300 TGGTGGATGCCACCTAACGATGCCAGAAAGTCTCCAACTCTCTATATAAGGAGCCTTGT 359

QY 360 ATTCAGTTGCAAAACAGCACCGACCAACCGGAGTTTCTCTGATTTGAGAAATATAAAACT 419

```

Db      360 ATTCAGTTGGAACACGACACACGAGTTACTCTCTGATTGAGAAATAAACT 419
QY      420 TCTGCTCTTGAACACACTTTGGGAGTTCA 451
Db      420 TCTGCTCTTGAACACACTTTGGGAGTTCA 451

RESULT 3
AA06864
ID      AA06864 standard; cDNA; 1322 BP.
XX
AC      AA06864;
XX
XX      26-APR-1999 (first entry)
XX
DE      Australian banana cv. Williams-infected badnavirus promoter pCv.
XX
KW      Promoter pCv; transgenic plant; banana; ds.
XX
OS      Badnavirus.
XX
FH      Key Location/Qualifiers
CDS     1..849
FT      /tag= a
FT      /note= "badnavirus ORF3 partial coding region"
FT      850..1322
FT      /tag= b
FT      /note= "this region is specifically claimed in
FT      Claim 2"
FT      1923..1322
FT      /tag= c
FT      /note= "this region is specifically claimed in
FT      Claim 3"
FT      protein_bind 997..1006
FT      /tag= d
FT      /label= GATA-1
FT      /note= "GATA binding factor 1 binding site"
FT      1073..1086
FT      /tag= e
FT      /label= AFF
FT      /note= "activation transcription factor binding
FT      site"
FT      TATA_signal 1135..1149
FT      /tag= f
FT      prim_transcript 1173..1181
FT      /tag= g
FT      /note= "transcription initiation site"
XX
FN      W09900492-A1.
XX
XX      07-JAN-1999.
XX
XX      26-JUN-1998; 98WO-AU00493.
XX
XX      26-JUN-1997; 97AU-0007593.
XX
XX      (SUGA-) BUREAU SUGAR EXPERIMENT STATIONS.
XX      (CSIR) COMMONWEALTH SCI & IND RES ORG.
XX      (QUEE-) STATE QUEENSLAND DEPT PRIMARY IND.
XX      (UYLE-) UNIV KATHOLIEKE LEUVEN.
XX      (UYQU) UNIV QUEENSLAND.
XX      (UYQU-) UNIV QUEENSLAND TECHNOLOGY.
XX
XX      Dietzgen RG, Elliott AR, Geering ADW, Grof CPL;
XX      P. McMichael LA, Remy S, Sagi L, Schenk PMP, Swennen RL;
XX      Thomas JE;
XX
XX      WPT; 1999-095738/08.
XX
XX      New promoter that is operable in a plant cell - useful in genetic
XX      engineering for regulation of gene expression
XX

```

```

PS      Claim 1; Page 26; 52pp; English.
XX
XX      this is the nucleotide sequence of the badnavirus promoter pCv
XX      that can be used to confer high-level gene expression on transgenic
XX      plants. The promoter was identified in viral DNA isolated from
XX      badnavirus-infected leaf material of Australian banana cv. Williams
XX      (Musa group AAA). Promoter DNA was obtained from viral DNA by PCR
XX      using degenerate primers badnat and badna3 (see AA06867-68).
XX      Putative promoter elements were identified using a computer program
XX      and by comparison with putative promoter elements of other plant
XX      virus genome promoters. Claimed promoters (see also AA06863 and
XX      AA06865) are useful for expressing a gene product in a plant cell,
XX      including a monocot such as sugarcane, banana, maize, millet or
XX      sorghum, a dicot such as tobacco, canola, tipu tree or Nicotiana
XX      benthamiana, a gymnosperm such as radiata pine, or a fern (all
XX      claimed). The gene products can confer e.g. disease resistance,
XX      herbicide resistance, improved tolerance to environmental factors,
XX      or can modulate plant composition, development, and fruit or crop
XX      quality.
XX
SQ      Sequence 1322 BP; 441 A; 239 C; 335 G; 307 T; 0 other;

Query Match      24.6%; Score 110.8; DB 20; Length 1322;
Best Local Similarity 67.4%; Pred. No. 1.7e-25;
Matches 260; Conservative 0; Mismatches 117; Indels 9; Gaps 7;

QY      60 GAAGCGGAAGTGGCGGACCCC-TACCACGTGTGTGATACCAACCGGTGTGA-AGACTGATA 117
Db      858 GTAGCGGAAGTGGTGGACCCCATACCACTGGATGGCACTAACCACTGTGACAAAGATACG 917
QY      118 AGATGGGAGTGGTGGATACCACTCTTATCTTAAGAGGAGACAAATATATGTC 177
Db      918 AGATGGCAAGTGGTGGATACCACTCTTATCTTAAGAGTGGTGTGCGTACCAACTC 977
QY      178 TCTTTATTTTAAAGTTTGTGGTGTGTTCTCTTAGTCACGACGATGACCTTTAGTGAAC 237
Db      978 ---CACTATAGTCTGTCTGAGGTGGATGCTGTGTGACGACAAAGA-CTTAGATTCCT 1033
QY      238 TTGCAGATCTTACGCAAGTGTGTAGCCAGACATGTGATGATCTTATCTGCATT 297
Db      1034 TTGCGTGAGATGTACCAAGCAGTGTGTCAGAGTGTGTGACGGTGTGCGTCCCTTGCATT 1093
QY      298 ATTGGTGGATGCCACCTAACGATGCCAAGACCTCCACAACCTCTCTATATAGGAGCCTT 357
Db      1094 ATTGGTGGGTG-CACCTAACGATGGGGAAGCGGAACCTCCCTCTATAAATAGGACCCGT 1152
QY      358 GTATTTCAGTTGCAACAGCCACCAACGCGAGTTACTCTCTGATTTGAGAAATAAAA 417
Db      1153 GTATTCA-GTTCAGACGACCAACACACGCGAGCTTACTTCTGAGAAGAAATAGAAC 1211
QY      418 CTTCTGTGCTTGAACACACACTTTGTG 443
Db      1212 ATT-TGTGCTTGAATACACCTTTGTG 1236

RESULT 4
AA05372/c
ID      AA05372 standard; DNA; 6071 BP.
XX
XX      AA05372;
XX
XX      18-DEC-2001 (first entry)
XX
XX      Chemically pretreated genomic DNA associated with cell cycle #39.
XX
XX      Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
XX      human immunodeficiency virus; neurodegenerative disorder; solid tumour;
XX      graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
XX      arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
XX      immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; OS;
XX      PCR primer.
XX
XX      Homo sapiens.
OS

```


Db 4058 TAATTCTAAAAAAGCTTTATATACTAAACACCTTCCACAAAAATTCGACTAAACCTCCT 3999
 QY 335 CAACCTCTCTATATAGGAGCCTTCTATTTCAGGTTGCACACACGACACACGCGAGTTT 394
 Db 3998 TAACCTCACTAATCACTACCTTAACCTCGATTTTCACCAAAACCTTAACCTCAATTT 3939
 QY 395 ACTCCTGATTTGAGAAATAAAA 417
 Db 3938 TCAAGCCTATAATAAAAAATAAA 3916

RESULT 6
 ID ABL32060 standard; DNA; 7516 BP.
 AC ABL32060;
 DT 26-MAR-2002 (first entry)
 DE Human immune system associated gene SEQ ID NO: 33.
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytostatic; neutrotropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 OS Homo sapiens.
 PN WO200200928-A2.
 PD 03-JAN-2002.
 XX 02-JUL-2001; 2001WO-EP07537.
 XX 30-JUN-2000; 2000DE-1032529.
 XX 01-SEP-2000; 2000DE-1043826.
 XX (EPIG-) EPIGENOMICS AG.
 PI Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-130909/17.
 XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 PS Claim 1; SEQ ID NO 33; 32pp + Sequence Listing; German.
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX Sequence 7516 BP; 2495 A; 36 C; 1376 G; 3609 T; 0 other;
 SQ Query Match 7.3%; Score 33; DB 24; Length 7516;
 Best Local Similarity 50.3%; Pred. No. 6.6; Indels 0; Gaps 0;
 Matches 81; Conservative 0; Mismatches 80;

QY 147 TTTATGTAAGAGGAGACAAAGTATATGCTCTTTATTTTAACTTTCGCTGCTGTTG 206
 Db 5323 TTTATATTAAATAAATTTATAGTAAATGATAGTTTATTTTAACTTATGATGATT 5382
 QY 207 TCTAGTCAGCAGGATGACCTTTAGTGACCTTCAGGATTTCTACCAAGTTGTTAGG 266

Db 5383 TGTAATTTAAAGTTAGAAAAGTTTTTAATGGTAAGTTATTTTATAAAGAAATATA 5442
 QY 267 CCAGAGCATGTGATGCTTATCTCCTCATTTATGGTGGAT 307
 Db 5443 ATTTAAATGCTTATTTATTTTAAATTTTAAATTTTATGAAT 5483

RESULT 7
 AAC79035/C
 ID AAC79035 standard; DNA; 1045 BP.
 XX AAC79035;
 AC AAC79035;
 DT 14-FEB-2001 (first entry)
 DE Human secreted protein gene 39 clone HPMB191.
 KW Cytostatic; immunosuppressive; neutrotropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein; ss.
 OS Homo sapiens.
 XX WO2000058358-A1.
 PN 05-OCT-2000.
 PD 23-MAR-2000; 2000WO-US07725.
 XX 26-MAR-1999; 99US-0126602.
 PR 14-JAN-2000; 2000US-0176063.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Ruben SM, Komatsoulis G;
 PI WPI; 2000-594640/56.
 DR P-PSDB; AAB44372.
 XX Forty nine nucleic acid molecules encoding human secreted proteins,
 PT useful in the prevention, treatment and diagnosis of cancer, immune
 PT disorders, cardiovascular disorders and neurological diseases -
 PS Claim 1; Page 333; 367pp; English.
 XX The invention relates to the isolation of genes AAC78997-C79045 encoding
 CC 49 human secreted proteins AAB44335-B44382. The genes can be used to
 CC generate fusion proteins by linking to the gene for the human
 CC immunoglobulin G Fc portion (AAC78988) for increasing the stability of
 CC the fusion protein as compared to the human protein only. The genes and
 CC proteins are useful for preventing, ameliorating or treating medical
 CC conditions, e.g. by protein or gene therapy. The genes are isolated
 CC from a range of human tissues disclosed in the specification. The
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
 CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 XX Sequence 1045 BP; 300 A; 204 C; 254 G; 287 T; 0 other;
 SQ Query Match 7.3%; Score 32.8; DB 21; Length 1045;
 Best Local Similarity 59.8%; Pred. No. 3.2;
 Matches 55; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 148 TTATGTAAGAGAGACAAAGTATATGTCCTCTTTATTTTAAAGTTTTCGGTGGTTGT 207
 Db 974 TTGTGACATAGGAGACAGATGTTATTTCTCTTTTACCAATACAAAATGTTATTTT 915
 QY 208 CTAGTCACGACGATGACCTTTAGTGAACCTT 239
 Db 914 CTGACCATGAACAATGACCTTCTGCAGTT 883

RESULT 8

ABL24210/c
 ID ABL24210 standard; DNA; 2257 BP.

XX AC ABL24210;
 XX DT

XX *26-MAR-2002 (first entry)

XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 24103.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical; gene; ds.

XX OS Drosophila melanogaster.

XX FN WO200171042-A2.

XX PR 27-SEP-2001.

XX XX 23-MAR-2001; 2001WO-US09231.

XX XX 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX XX (PEKE) PE CORP NY.

XX FI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

XX PS Claim 1; SEQ ID NO 24103; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB5737-ABB72072).

XX CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pot_sequences.

XX SQ Sequence 2257 BP; 611 A; 455 C; 543 G; 648 T; 0 other;

Query Match 7.3%; Score 32.8; DB 23; Length 2257;
 Best Local Similarity 55.2%; Pred. No. 4.5;
 Matches 64; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 174 TGCTCTTATTTAAAGTTTTCGGTGTCTAGTCACGACGATGACCTTTAGTG 233

Db 1717 TTTTCTTTTGTTCGTGTTGCGGTGCTTTGCGCTTTTCACACGCCACTATTAAAC 1658

QY 234 AACTTTGCAGGATCTTCAGCAAGTTGTTAGCCAGACAGATGATGATGCTTA 289

Db 1657 CCAGAGTGGCGATGAGTACTGAAATATTATTTCAAACATATGACTGATCAAA 1602

RESULT 9

ABV14414

XX ID ABV14414 standard; cDNA; 654 BP.

XX AC ABV14414;

XX DT 13-SEP-2002 (first entry)

XX DE Human prostate expression marker cDNA 14405.

XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW KW pharmacogenomic marker; gene; ss.
 OS Homo sapiens.
 XX FN WO200160860-A2.
 XX PR 23-AUG-2001.
 XX XX 20-FEB-2001; 2001WO-US05171.
 XX PR 17-FEB-2000; 2000US-183319P.
 XX PR 16-MAR-2000; 2000US-189862P.
 XX PR 23-MAY-2000; 2000US-207454P.
 XX PR 03-JUN-2000; 2000US-211314P.
 XX PR 18-JUL-2000; 2000US-219007P.
 XX PR 13-DEC-2000; 2000US-255281P.
 XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX PI Schlegel R, Endege WO, Monahan JE;
 XX DR WPI; 2001-662795/76.

XX PT Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -

XX PS Claim 1; Page 2405; 11750pp; English.

XX CC The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213), of the specification or its complement. (I) is useful for:

XX CC (a) assessing whether a patient is afflicted with prostate cancer;

XX CC (b) monitoring the progression of prostate cancer in a patient;

XX CC (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;

XX CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;

XX CC (e) selecting a composition for inhibiting prostate cancer in a patient;

XX CC (f) assessing the prostate cell carcinogenic potential of a compound;

XX CC (g) determining whether prostate cancer has metastasized in a patient;

XX CC (h) assessing the aggressiveness or indolence of prostate cancer in a patient;

XX CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX SQ Sequence 654 BP; 120 A; 148 C; 194 G; 187 T; 5 other;

Query Match 7.1%; Score 32.2; DB 23; Length 654;
 Best Local Similarity 54.2%; Pred. No. 4.1;
 Matches 64; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 136 ATACACTCAGTTTATGTAAGAGAGAGAGATAATGTCCTTTATTTTAAAGTTGT 195

Db 402 ATGCAAGGCCCTTTGTTATTCGGAGTGATCGCGGTTTCGGATCTCTTCTCAGAGCGGT 461

QY 196 CGGTGTCGTCTAGTCACGACCATGACCTTAGTGAACCTTCAGGATCTTACG 253

Db 462 GGGTGTGCTGTGTGTACCCCGCCCTTTTGTGTTTTCCTTTTAAAG 519

RESULT 10

ABQ53082


```

Db 353 TATATTTTAAATTTGCGTTTATTATGGAAGTTTCGTTTATTGTTAAGTTTAC 294
Qy 237 TTT 239
Db 293 GTT 291

RESULT 12
AA14043/c
ID AAX14043 standard; DNA; 1180 BP.
XX
AC AAX14043;
XX
DT 31-MAR-1999 (first entry)
DE H. pylori GHPO 1411 gene.
XX
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease; ss.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT CDS 51..1130
FT FT /*tag= a
XX
PN WO9843478-A1.
PD 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-US06371.
XX
PR 29-JUL-1997; 97US-0902615.
PR 01-APR-1997; 97US-0833457.
PR 24-JUN-1997; 97US-0881227.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX
DR WPI; 1998-542293/46.
DR P-PSDB; AAW98324.
XX
XX New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
XX
PS Claim 1; Page 435-437; 2054pp; English.
XX
XX This sequence represents a polynucleotide of the invention. It was
CC isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
CC The polypeptides can be used for preventing or treating Helicobacter
CC infections, and gastroduodenal diseases associated with these
CC infections, including acute, chronic, and atrophic gastritis, and peptic
CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
SQ Sequence 1180 BP; 404 A; 171 C; 296 G; 309 T; 0 other;

Query Match 7.1%; Score 31.8; DB 19; Length 1180;
Best Local Similarity 50.3%; Pred. No. 7.2;
Matches 78; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 139 CCATCTATTATGTAAGGAGACAACTATTAATGCTTTATTATTAAGTTTTCGG 198
Db 378 CAACCACTTTAAATAGTAAGCAACAACTTTGCCAACTGATTTTATTAGGATCATAA 319
Qy 199 TGTCGTTGTCAGTCAGCGATGACCTTTAGTGAACCTTTGCAGGATTTCTTACGCAAG 258
Db 318 TGACTTTCCAGCTTTCAGCATGATCACTTTTCATGATTTTCTCATTAATTCGTTGAG 259

```

```

Qy 259 TTGTAGGCGACAGACATGTGATGATGCTTATCTG 293
Db 258 TCTTGCGGTAGCGTAACACAGAGCTTGCCTATG 224

RESULT 13
AAV31267/c
ID AAV31267 standard; DNA; 1234 BP.
XX
AC AAV31267;
XX
DT 01-OCT-1998 (first entry)
DE E. coli J96 pathogenicity island contig #81.
XX
KW PAI; pathogenicity island; uropathogenic E. coli detection; PAI IV; pher;
KW PAI V; pher; vaccine; protective immune response; ds.
XX
OS Escherichia coli.
XX
PN WO9822575-A2.
PD 28-MAY-1998.
XX
PF 21-NOV-1997; 97WO-US21347.
XX
PR 14-OCT-1997; 97US-0061953.
PR 22-NOV-1996; 96US-0031626.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (UYWI-) UNIV WISCONSIN.
XX
PI Choi GH, Dillon PJ, Welch RA;
XX
DR WPI; 1998-312461/27.
XX
PT New isolated uropathogenic E. coli nucleotide sequences - used to
PT develop products for the detection of pathogenic E. coli and to
PT elicit an immune response to pathogenic E. coli
XX
PS Claim 21; Page 196; 250pp; English.
XX
XX This sequence represents a E. coli strain J96 contig containing
CC pathogenicity island (PAI) sequences, and represents a nucleic acid
CC molecule of the invention. PAIs are large fragments of DNA which comprise
CC pathogenicity determinants. The sequences of the invention are taken from
CC PAI IV and PAI V. PAI IV is located at approximately 64 min (near pher)
CC on the E. coli chromosome and is greater than 170 kb. PAI V is located at
CC approximately 94 min (at pher) on the E. coli chromosome and is
CC approximately 160 kb in size. Antibodies specific to the proteins encoded
CC by the PAI open reading frames of the invention can be used in kits to
CC detect uropathogenic E. coli. The proteins are used in vaccines to elicit
CC a protective immune response in an animal to the uropathogenic E. coli
XX strain J96.
XX
SQ Sequence 1234 BP; 343 A; 250 C; 261 G; 377 T; 3 other;

Query Match 7.1%; Score 31.8; DB 19; Length 1234;
Best Local Similarity 76.5%; Pred. No. 7.3;
Matches 39; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 135 GATACCACTCATTATGTAAAGAGGAGACAAAGTATATGTCCTTTATT 185
Db 789 GATACCAATCTCTTACAGAAAGAGTCGACAAAGGAATTCGCTTCATT 739

RESULT 14
AAS56470
ID AAS56470 standard; CDNA; 214 BP.
XX
AC AAS56470;
XX

```


DT	18-DEC-2001	(first entry)
XX		
DE	Human cDNA for an ovarian cancer protein #94.	
XX		
KW	Human; ss; ovarian cancer protein; cancer; tumour; ovarian cancer;	
KW	endometrial cancer; cytostatic.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200170976-A2.	
XX		
PD	27-SEP-2001.	
XX		
PF	20-MAR-2001; 2001WO-US09062.	
XX		
PR	21-MAR-2000; 2000US-190710P.	
PR	22-JUN-2000; 2000US-213748P.	
PR	19-DEC-2000; 2000US-257276P.	
XX		
XX	(CORI-) CORIXA CORP.	
PA		
PI	Xu J, Pyle RA, Stolk JA;	
XX		
DR	WPI; 2001-607531/69.	
XX		
PT	Nucleic acids encoding 222 polypeptides associated with ovarian and endometrial cancers, useful for diagnosing, preventing and treating cancers -	
PT		
PS	Claim 1; Page 146-147; 187pp; English.	
XX		
CC	The invention relates to human polynucleotides encoding proteins associated with ovarian and endometrial cancers. The polynucleotides and the proteins they encode may be used in the prevention, diagnosis and treatment of diseases associated with the inappropriate expression of ovarian and endometrial cancer polypeptides (OECPs). For example, the polynucleotide (or an expression vector comprising the polynucleotide) and the OECp may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of OECps by expressing inactive proteins or to supplement the patients own production of them. Additionally, the polynucleotide may be used to produce the OECps, by inserting the nucleic acids into a host cell and culturing the cell to express the protein. The polynucleotide and its complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acids in samples, and therefore which patients may be in need of restorative therapy. The OECps may also be used as antigens in the production of anti-OECP antibodies and in assays to identify modulators of it's expression and activity. The anti-OECP antibodies and antagonists may also be used to down regulate expression and activity. The anti-OECP antibodies may also be used as diagnostic agents for detecting the presence of OECP in samples (e.g. by enzyme linked immunosorbent assay (ELISA)) and hence diagnose patients with cancers. The present sequence is a ovarian and endometrial cancer linked cDNA of the invention.	
XX		
SQ	Sequence 214 BP; 64 A; 34 C; 23 G; 90 T; 3 other;	
	Query Match	7.0%; Score 31.6; DB 22; Length 214;
	Best Local Similarity	51.5%; Pred. No. 4;
	Matches	70; Conservative 0; Mismatches 56; Indels 0; Gaps 0
QY	178 TCCTTTATTTTAAGTTCGCGTGTCGTCTAGTCACGCAGCATGCCCTTTAGTCAACT	237
Dd		
	4 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNGCAACAACAGTCAATCTCTATTGGAAA	63
QY	238 TTGCGAGGATCTTAGCGAAATCTGTAGGCCAGAGACATGTGATGATGCTTATCGCATT	297
Dd		
	64 CTGCGAGTATTATAATACATACAAATCTCTGTGTACATAAACGTCGCTTTGANATTTTAAAT	123
QY	298 ATTGGTGGATGCCACC	313
Dd		
	124 CTGAGCTCATCTCATC	139

Db 296 ATTTCCTTAATACATTTCCTTAATC 270

Search completed: June 24, 2003, 04:50:48
Job time : 191 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 03:09:20 ; Search time 1128 Seconds
(without alignments)
6475.323 Million cell updates/sec

Title: US-09-622-978-2

Perfect score: 451

Sequence: 1 agacaagaatatcttattg.....acacactttgtgcgagtcca 451

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_estham:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estc:*
9: gb_esti:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_othr:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	38.8	8.6	525	10	AV563822
c 2	37.6	8.3	531	10	BE203321
c 3	37.6	8.3	1019	17	CNS00JFW
4	37	8.2	615	17	AQ646390
c 5	36.8	8.2	538	17	AZ313225
6	36.4	8.1	527	17	TA57A01Q

7	36.4	8.1	604	17	TA198A02P
8	36.4	8.1	702	17	AQ948036
9	35.8	7.9	418	17	CNS030P8
c 10	35.8	7.9	551	17	AQ176795
11	35.4	7.8	640	17	A3977903
12	35.2	7.8	925	12	BG104629
13	35	7.8	344	10	BB330407
14	34.6	7.7	537	13	BM318785
15	34.6	7.7	540	10	AW677574
16	34.6	7.7	633	14	B0480977
c 17	34.4	7.6	302	9	AL380877
18	34.4	7.6	393	13	BI023396
c 19	34.4	7.6	846	12	BG585827
20	34.4	7.6	848	17	CNS02ICR
c 21	33.8	7.5	610	17	A2737773
22	33.6	7.5	978	17	CNS014NO
c 23	33.4	7.4	420	10	AW970814
c 24	33.4	7.4	490	9	AA482025
c 25	33.4	7.4	612	12	BG593863
c 26	33.4	7.4	742	12	BE865954
c 27	33.4	7.4	884	13	BI914220
28	33.2	7.4	731	17	BH601473
29	33.2	7.4	845	17	CNS03NZ2
30	33.2	7.4	888	13	BI914880
c 31	33	7.3	168	12	BF837389
c 32	33	7.3	450	12	BG603483
c 33	33	7.3	453	14	W58344
c 34	33	7.3	545	17	W58344
35	33	7.3	625	10	BI153800
c 36	33	7.3	626	17	CNS030CW
c 37	33	7.3	738	17	A2347718
c 38	33	7.3	743	12	BE870824
c 39	33	7.3	1101	17	CNS0029D
c 40	32.8	7.3	308	9	AU268299
41	32.8	7.3	369	9	AI381477
42	32.8	7.3	382	9	AI184653
43	32.8	7.3	467	10	AW802848
44	32.8	7.3	561	13	BM093002
c 45	32.8	7.3	992	17	CNS029LD

ALIGNMENTS

RESULT 1
AV563822/c
LOCUS AV563822 525 bp mRNA linear EST 07-SEP-2000
DEFINITION Arabidopsis thaliana green siliques Columbia Arabidopsis
thaliana cDNA clone SQ194a08f 3', mRNA sequence.

ACCESSION AV563822
VERSION AV563822.1 GI:8735248
KEYWORDS EST.

SOURCE
thale cress.

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

AUTHORS

TITLE

Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries

JOURNAL

MEDLINE

COMMENT

Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES

source

1. 525

/organism="Arabidopsis thaliana"

/strain="Columbia"

```

/db_xref="taxon:3702"
/clone="SQ194a08F"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 181 a 95 c 114 g 135 t
ORIGIN

Query Match 8.6%; Score 38.8; DB 10; Length 525;
Best Local Similarity 45.9%; Pred No. 0.21;
Matches 133; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 135 GATACCACCTCATTTATGTAAGAGAGACAAAGTATATGCTCTTTATTTTAAAGTTG 194
Db 291 GTTTCGGGTGACCTCTGTTTCTTAGCATCAAGAACAAGTCTCTTTTCTGTTTTT 232
QY 195 TCGGTGCTGTGTCATGTCACGACGATGACCTTTAGTGAAGTTTGCAGGATTTTACGC 254
Db 231 ACCGAAGTTCGAAGTCCCTTTTATAGTTAGTTAGTGTGATGTCATTTATAGC 172
QY 255 AAGTTGTTAGCCGACGACATGTTGATGATGCTTATCTGCATTTATGTTGATGCCACCT 314
Db 171 TTGGTGTGTTTTCAGAAAGTCAGTCAAGTTTCTCTCTCTTTTATAGTTTCCACCA 112
QY 315 AAGCATGCCAGAAAGCTCCCAACCTCTCTATATAAGAGCCCTTGATTCAGTTGCAAC 374
Db 111 GAAATAATTAATCTTGTGAAGTGTGTTTGAAGTATGTTCTTTATGTTTTCACAC 52
QY 375 ACGCACCAACACGCGAGTTTACTCTCTGATTTGAGAATAAAGCTTCTGT 424
Db 51 CGAAAGAACAATATTTTCTTTGTTGTAATTAATAAGATTTACTTCTGT 2

RESULT 2
BE203321/c 531 bp mRNA linear EST 07-SEP-2000
LOCUS BE203321
DEFINITION BE203321 KVI Medicago truncatula cDNA clone PKV1-5K23, mRNA
sequence.
ACCESSION BE203321
VERSION BE203321.1 GI:8746592
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 531)
AUTHORS VandenBosch,K., Endre,G., Hur,J., Moore,J., Beremand,P., Ellis,L.,
Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and
Fraser,C.M.
ESTS from roots of Medicago truncatula 24 hours after inoculation
with Sinorhizobium meliloti
UNPUBLISHED (1999)
CONTACT: VandenBosch K
Department of Plant Biology
University of Minnesota
220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738
Email: kvandenb@csb.umn.edu
Texas A&M University: T263093e
TIGR sequence name: MTIAM72TK
More information is available at:
http://chrystle.tamu.edu/medicago
Seq primer: Skmcd (CTA GAA CTA gtg gat CC).
Location/Qualifiers
1. 531
/organism="Medicago truncatula"
/cultivar="genotype AL7"
/db_xref="taxon:3880"
/clone="pkv1-5K23"

FEATURES
source

```

```

/clone_lib="KVI"
/tissue_type="Seedling roots"
/dev_stage="24 hours post-inoculation with Sinorhizobium meliloti"
/lab_host="E. coli strain XL0LR"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XL0LR cells."
BASE COUNT 157 a 80 c 130 g 184 t
ORIGIN

Query Match 8.3%; Score 37.6; DB 10; Length 531;
Best Local Similarity 61.0%; Pred. No. 0.53;
Matches 61; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 306 ATGCCACCTAAGATGCCAGAAAGCTCCACAACCTCTCTATATAAGAGCCCTTGATTCAG 365
Db 111 ATGTCCCATTCCTCAATCTCCACAGCACTGCAACTTGCAGCAATTATTAACCTGATTTTCAG 52
QY 366 GTTGCAAAACACGACCAACCAACGCGAGTTTACTCTCTGATTT 405
Db 51 ATTCAAACACGATCCCAAAATCGAATTAGTTCAGATTT 12

RESULT 3
CNS000JPW 1019 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC:
DEFINITION BACR38L16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL076789
VERSION AL076789.1 GI:4956165
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1019)
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 131 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1. 1019
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR38L16"
/clone_lib="RPCI-98"
/notes="end : TET3"

BASE COUNT 243 a 247 c 254 g 218 t 57 others
ORIGIN

```

Query Match	8.28;	Score 37;	DB 17;	Length 615;
Best Local Similarity	50.0%;	Pred. No. 0.89;		
Matches 91;	Conservative 0;	Mismatches 91;	Indels 0;	Gaps 0;
QY	101	CGGTGTGAAGACTGATAGATGCGGAGTGAAGCTGGATACCACTCACTTATGTATTAAGAGG	160	
DB	403	CGGTGTGCACAGTGNAGTGTGACAGTGAATTAGATAGCATTCGCCAGCGAAGCGCGT	462	
QY	161	AGACAAAGTATAATGTCTCTCTTATTTTAAAGTTGTTCGGTGTCTGTCTAGTCCACCGAG	220	
DB	463	CTACTAGGAAGATTATTTTATTTTCTTTTGTTCACGCGCGGCATTTAGTGAGAGATG	522	
QY	221	ATGACCTTTAGTGAACCTTTGCAGGATCTTACGCAAAAGTTGTTAGGCCAGAGACATGTGA	280	
DB	523	TCCGGATGTGCGTTAATAAGGAAGCGTACGAATATTTTCATANGTACATTTTCATGTCA	582	
QY	281	TG 282		
DB	583	AG 584		

RESULT 5						GSS 29-SEP-2000
LOCUS	AZ313225/c					
DEFINITION	mouse 10kb plasmid UUGCIM library Mus musculus genomic clone UUGCIM0029d18 R, DNA sequence.					
ACCESSION	AZ313225	538 bp	DNA	linear		
VERSION	AZ313225.1					
KEYWORDS	GSS.					
SOURCE	house mouse.					
ORGANISM	Mus musculus					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
REFERENCE	1 (bases 1 to 538)					
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Stokes,R., Tingey,A., von Niederhausern,A., Wright,D., Weiss,R.					
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts					
JOURNAL	Unpublished (2000)					
COMMENT	Contact: Robert B. Weiss					
	University of Utah Genome Center					
	University of Utah					
	Rm. 308; Biomedical					
	Polymer Science Research Bldg.,					
	20 S. 2030 E., SLCT,					
	UT,					

```

Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 538.
Location/Qualifiers
    i. 538
    /organism="Mus musculus"
    /strain="C57BL/6J"
    /db_xref="taxon:10090"
    /clone="UUGC1M0029D18"
    /clone_lib="Mouse 10kb plasmid UUGC1M library"
    /sex="Male"
    /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
    /note="Vector: PWD42nv; Purified genomic DNA from M.
    musculus C57BL/6J (male) was obtained from the Jackson
    Laboratory Mouse DNA Resource
    (http://www.jax.org/resources/documents/dnares/). The DNA
    was hydrodynamically sheared by repeated passage through a
    0.005 inch orifice at constant velocity. The sheared DNA
    was blunt end-repaired with T4 DNA polymerase and T4
    polynucleotide kinase. Adaptor oligonucleotides were

```

	Query Match	8.1%	Score 36.4;	DB 17;	Length 604;
	Best Local Similarity	50.0%;	Pred. No. 1.4;	Mismatches	91; Indels 0; Gaps 0;
	Matches	91; Conservative	0; Mismatches	91;	Indels 0; Gaps 0;
QY	101	CGGTGTTGAAGACTCATAGATCGGGACGTGACGTGCATACCCTCACTTATTCATAAAGAGG	160		
Dd	103	CGGTGTGCACAGTCAGATGTTGACAAGTGAAATAGATAGCATTCCGCCACGAAAGCGGT	162		
QY	161	AGACAAAAGTAATGTCCTCTTTATTTTAAAGTTTGTGGTGTCTGTCATGTCAGTCAGCACG	220		

```

Db 163 CTACTAGAAAGTATATTTTATTTTCTTTTCTTTGTTGACGGCGCATTTTACTGAGAGATG 222
QY 221 ATGACCTTAGTGAACCTTTGAGGATCTTTACGCAAAAGTGTGTAGGCCAGACATGTGA 280
Db 223 TCGCGATGCGGTAAATAAGGAAGAGCGTACGAATATTTTCATAGTTACATTTTCATGTCA 282
QY 281 TG 282
Db 283 GG 284

RESULT 8
LOCUS AQ948036
DEFINITION Sheared DNA-45P20, TF Sheared DNA Trypanosoma brucei genomic clone
ACCESSION AQ948036
VERSION AQ948036.1 GI:6771301
SOURCE GSS.
ORGANISM Trypanosoma brucei.
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 702)
El-Sayed,N., Zhao,S., Zhao,H., Gall,S., Suh,E., Malek,J., Fujii,C.,
Gerrard,C., Leech,V., de Jong,P., Ulliu,E., Melville,S., Donelson,J.,
Fraser,C. and Adams,M.
Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library
Unpublished (1999)
JOURNAL Other GSSs: Sheared DNA-45P20.TR
COMMENT Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through ATCC. Sheared DNA end sequences search page:
http://www.tigr.org/tldb/mbd/tbdb/.
Seq primer: M13-Forward
Class: shotgun.
FEATURES
source
1..702
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone_lib="Sheared DNA-45P20"
/clone="Vector: pUC18; Site 1: SmaI; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + i method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. In Genome Sequencing: A Practical
Approach, eds. M. Vaudin and B. Barrell, Oxford University
Press, 1999)."
BASE COUNT 216 a 103 c 191 g 192 t
ORIGIN
Query Match 8.1%; Score 36.4; DB 17; Length 702;
Best Local Similarity 50.0%; Pred. No. 1.5;
Matches 91; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 101 CGGTGTGAAGACTGATAGATCGGAGTGGAGTGGATACCACTACTTATGTAAAGAGG 160
Db 26 CCGTGTGCACAGTGATGTGACAGTGAATTAGATAGCATTCGCCAGCGAGCGGT 85

```

```

QY 161 AGACAAAGTAAATGCTCTCTTTATTTTAAAGTTTCTCGGTGCTGTTCTAGTACGCAAG 220
Db 86 CTACTAGAAAGTATATTTTATTTTCTTTTCTTTGTTGACGGCGCATTTTACTGAGAGATG 145
QY 221 ATGACCTTTAGTGAACCTTTGAGGATCTTTACGCAAAAGTGTGTAGGCCAGACATGTGA 280
Db 146 TCGCGATGCGGTAAATAAGGAAGAGCGTACGAATATTTTCATAGTTACATTTTCATGTCA 205
QY 281 TG 282
Db 206 GG 207

RESULT 9
LOCUS CNS030P8
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
047L18 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL256229.1 GI:7977241
VERSION AL256229
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 418)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Berrot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
REFERENCE 2 (bases 1 to 418)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Ballault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
REFERENCE 3 (bases 1 to 418)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000)
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
Location/Qualifiers
1..418
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="047L18"
/clone_lib="G"
/note="Genoscope sequence ID : C0B3047DF09SP1-end :
PUC-ori"
BASE COUNT 108 a 67 c 89 g 135 t 19 others
ORIGIN
Query Match 7.9%; Score 35.8; DB 17; Length 418;
Best Local Similarity 51.8%; Pred. No. 1.9;
Matches 73; Conservative 3; Mismatches 65; Indels 0; Gaps 0;

QY 109 AGACTGATAGATCGGAGTGGAGTACCATCCACTTATGTATGATGATGAGAGAGACAAAG 168
Db 133 AGATCCAAAGATGAASATTCGATTCGACATTCGATTCGATTCGATTCGATTCGATTCG 192
QY 169 TATAATGTCCTTTATTTTAAAGTTTCTCGGTGCTGTTCTAGTACGCAAGACCTT 228
Db 193 TATAGTGTCTTTTCTTTTAAATATTGAGTTGTTCTTTCTAGTCTGTATYGGACCTT 252

```



```

QY 229 TAGTGAACCTTGCAGGATTCCT 249
DB 253 TTTMTATCATCAACAGTTCCT 273

RESULT 10
A0176795/c 551 bp DNA linear GSS 17-OCT-1998
LOCUS HS_3213_al_E12.T7 CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate-3213 Col-23 Row-I, DNA sequence.
ACCESSION A0176795
VERSION A0176795.1 GI:3574162
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 551)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 93980589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3213 row: I column: 23
Class: BAC ends
High quality sequence stop: 551.
Location/Qualifiers
1..551
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-3213 Col-23 Row-I"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 201 a 81 c 100 g 152 t 17 others
ORIGIN
Query Match 7.9%; Score 35.8; DB 17; Length 551;
Best Local Similarity 50.3%; Pred. No. 2.1;
Matches 85; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 143 TCACCTTATGTAAAGAGGACAAAGTATATGCTCTTTATTTAAAGTTTCTCGTGTC 202
DB 144 TTTTCTTATGTAAAGAGGACAAAGTATATGCTCTTTATTTAAAGTTTCTCGTGTC 202
328 TCACCTGATGTTAAAGTATCAAGAGAGAGAGGTGTCATGACACAGATATGTTATCTTAC 269

QY 203 GTTCTCTAGTCACGACGATGACCTTAGTGAACCTTTCGAGGATCTTACGCAAAAGTGT 262
DB 264 TTTTCTCTAGTCACGACGATGACCTTAGTGAACCTTTCGAGGATCTTACGCAAAAGTGT 262
268 CTGCAANAGTCAGCCTAGAGCATCTAGGCGCCCATCTGCTAGTATTGGGAAAGTTGT 209

QY 263 TAGCCGACAGACATGTCATGATCTTATCTGCATTTATTTGGTGATGCCA 311
DB 264 TTTTCTCTAGTCACGACGATGACCTTAGTGAACCTTTCGAGGATCTTACGCAAAAGTGT 262
208 AATCATATAGTTTATATGATCCAGATTTTAAATATATGTGAATCACA 150

RESULT 11
A2977903 640 bp DNA linear GSS 27-APR-2001
LOCUS 2M0254B01F Mouse 10kb plasmid UUGC2M library Mus musculus genomic
DEFINITION clone UUGC2M0254B01 F, DNA sequence.
ACCESSION A2977903
VERSION A2977903.1 GI:13849130
KEYWORDS GSS.

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 640)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D.,Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
JOURNAL
COMMENT Contact: Robert B. Weiss
University of Utah
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0254 row: B column: 01
Seq Primer: CGTTGTAACACGACGCCAGT
Class: plasmid ends
High quality sequence stop: 640.
Location/Qualifiers
1..640
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0254B01"
/clone_lib="Mouse 10kb plasmid UUGC2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (female) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (GI4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 231 a 105 c 130 g 174 t
ORIGIN
Query Match 7.8%; Score 35.4; DB 17; Length 640;
Best Local Similarity 66.2%; Pred. No. 3;
Matches 51; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 269 AGAGACATGTGATGATGCTTATCTGCATTTATTTGGTGGATGCCACCTACGATGCCAGAA 328
DB 207 AGAGAGATGGAGGTCATCTGATTCATAGCTAGTGGAGATGAAAGAGAAAGATGCCAGAA 266
329 GCTCCACCAACTCTCTAT 345
DB 267 GCTCCATATTTCTCTAT 283

RESULT 12
BG104629
LOCUS BG104629
DEFINITION 602311315F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4421187 5',

```


Search completed: June 24, 2003, 05:09:50
Job time : 1133 secs


```
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-081-320-24

Query Match
Best Local Similarity 6.7%; Score 30.4; DB 3; Length 447;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 48 GTAATGATGACTGAAGCGGAGAGTGGCGGACCCCTACCAACGTTGTGATACCAACCGGTGTG 107
Db 361 GAAATGATGACAGAAGAAGAGAGTGGCGGACACACTACATTTGTGTGAGGTTCTCGTTAGG 420
QY 108 AAGACTGATAAGATCGGAG 127
Db 421 AACAAGCATAGATGAGGTG 440

RESULT 5
US-09-574-141A-24
; Sequence 24, Application US/09574141A
; Patent No. 6395490
; GENERAL INFORMATION:
; APPLICANT: Gonsalves, Dennis
; APPLICANT: Meng, Baozhong
; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES
; FILE REFERENCE: 07678/035005
; CURRENT APPLICATION NUMBER: US/09/574,141A
; CURRENT FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: 60/047,147
; PRIOR FILING DATE: 1997-05-20
; PRIOR APPLICATION NUMBER: 60/069,902
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 09/081,320
; PRIOR FILING DATE: 1998-05-19
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Rupestris stem pitting associated virus
US-09-574-141A-24

Query Match
Best Local Similarity 6.7%; Score 30.4; DB 4; Length 447;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 48 GTAATGATGACTGAAGCGGAGAGTGGCGGACCCCTACCAACGTTGTGATACCAACCGGTGTG 107
Db 361 GAAATGATGACAGAAGAAGAGAGTGGCGGACACACTACATTTGTGTGAGGTTCTCGTTAGG 420
QY 108 AAGACTGATAAGATCGGAG 127
Db 421 AACAAGCATAGATGAGGTG 440

RESULT 6
US-09-707-780-24
; Sequence 24, Application US/09707780
; Patent No. 6399308
; GENERAL INFORMATION:
; APPLICANT: Gonsalves, Dennis
; APPLICANT: Meng, Baozhong
; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES
; FILE REFERENCE: 07678/035006
; CURRENT APPLICATION NUMBER: US/09/707,780
; CURRENT FILING DATE: 2000-11-07
; PRIOR APPLICATION NUMBER: 09/081,320
; PRIOR FILING DATE: 1998-05-19
; PRIOR APPLICATION NUMBER: 60/047,147
```

```
; PRIOR FILING DATE: 1997-05-20
; PRIOR APPLICATION NUMBER: 60/069,902
; PRIOR FILING DATE: 1997-12-17
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Rupestris stem pitting associated virus
US-09-707-780-24

Query Match
Best Local Similarity 6.7%; Score 30.4; DB 4; Length 447;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 48 GTAATGATGACTGAAGCGGAGAGTGGCGGACCCCTACCAACGTTGTGATACCAACCGGTGTG 107
Db 361 GAAATGATGACAGAAGAAGAGAGTGGCGGACACACTACATTTGTGTGAGGTTCTCGTTAGG 420
QY 108 AAGACTGATAAGATCGGAG 127
Db 421 AACAAGCATAGATGAGGTG 440

RESULT 7
US-09-081-320-23
; Sequence 23, Application US/09081320
; Patent No. 6093544
; GENERAL INFORMATION:
; APPLICANT: Gonsalves, Dennis
; APPLICANT: Meng, Baozhong
; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/081,320
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/047,147
; FILING DATE: 20-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/069,902
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1722
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2009 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-081-320-23

Query Match
Best Local Similarity 6.7%; Score 30.4; DB 3; Length 2009;
```

Best Local Similarity 61.2%; Pred. No. 1.9;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 48 GTAATGATGACTGAAGCGGAGTGGCGGACCCCTACCACTGTTGTATACCAACCGGTGTG 107
DB 361 GAAATGATGACAGAGAAGAGTGGAGGCACACTACATGTTGTGAGGTTCTGTGTAGG 420
QY 108 AAGACTGATAAGATGCGGAG 127
DB 421 AACAGCATAAGATGAGGTG 440

RESULT 8

US-09-574-141A-23
; Sequence 23, Application US/09574141A
; Patent No. 6395490
; GENERAL INFORMATION:
; APPLICANT: Gonsalves, Dennis
; APPLICANT: Meng, Baozhong
; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES
; FILE REFERENCE: 07678/035005
; CURRENT APPLICATION NUMBER: US/09/574,141A
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: 60/047,147
; PRIOR FILING DATE: 1997-05-20
; PRIOR APPLICATION NUMBER: 60/069,902
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 09/081,320
; PRIOR FILING DATE: 1998-05-19
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 2009
; TYPE: DNA
; ORGANISM: Rupestris stem pitting associated virus
US-09-574-141A-23

Query Match 6.7%; Score 30.4; DB 4; Length 2009;
Best Local Similarity 61.2%; Pred. No. 1.9;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 48 GTAATGATGACTGAAGCGGAGTGGCGGACCCCTACCACTGTTGTATACCAACCGGTGTG 107
DB 361 GAAATGATGACAGAGAAGAGTGGAGGCACACTACATGTTGTGAGGTTCTGTGTAGG 420
QY 108 AAGACTGATAAGATGCGGAG 127
DB 421 AACAGCATAAGATGAGGTG 440

RESULT 9

US-09-707-780-23
; Sequence 23, Application US/09707780
; Patent No. 6399308
; GENERAL INFORMATION:
; APPLICANT: Gonsalves, Dennis
; APPLICANT: Meng, Baozhong
; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES
; FILE REFERENCE: 07678/035006
; CURRENT APPLICATION NUMBER: US/09/707,780
; PRIOR FILING DATE: 2000-11-07
; PRIOR APPLICATION NUMBER: 09/081,320
; PRIOR FILING DATE: 1998-05-19
; PRIOR APPLICATION NUMBER: 60/047,147
; PRIOR FILING DATE: 1997-05-20
; PRIOR APPLICATION NUMBER: 60/069,902
; PRIOR FILING DATE: 1997-12-17
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 2009

; TYPE: DNA
; ORGANISM: Rupestris stem pitting associated virus
US-09-707-780-23
Query Match 6.7%; Score 30.4; DB 4; Length 2009;
Best Local Similarity 61.2%; Pred. No. 1.9;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 48 GTAATGATGACTGAAGCGGAGTGGCGGACCCCTACCACTGTTGTATACCAACCGGTGTG 107
DB 361 GAAATGATGACAGAGAAGAGTGGAGGCACACTACATGTTGTGAGGTTCTGTGTAGG 420
QY 108 AAGACTGATAAGATGCGGAG 127
DB 421 AACAGCATAAGATGAGGTG 440

RESULT 10

US-09-701-685-1
; Sequence 1, Application US/09701685
; Patent No. 6387629
; GENERAL INFORMATION:
; APPLICANT: Schneider, Patrick
; APPLICANT: Yamamoto, Karen K.
; APPLICANT: French, Cynthia K.
; APPLICANT: Reprogen, Inc.
; TITLE OF INVENTION: Use of Cathepsin S in the Diagnosis and Treatment of
; TITLE OF INVENTION: Endometriosis
; FILE REFERENCE: 018002-001310HS
; CURRENT APPLICATION NUMBER: US/09/701,685
; CURRENT FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: WO PCT/US99/12335
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: US 60/088,017
; PRIOR FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1643
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (70)..(1065)
; OTHER INFORMATION: human cathepsin S
US-09-701-685-1

Query Match 6.7%; Score 30.2; DB 4; Length 1643;
Best Local Similarity 50.3%; Pred. No. 2;
Matches 74; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
QY 155 AAGAGGAGACAAAGTATATGTCCTTTTAAAGTTTGGTGGTGTCTAGTCA 214
DB 1496 AAAAGAAAATATAGTTTGTATCTCTTTTAAATTTGCAAACTCAGGATAAGTTT 1555
QY 215 CGCAGCATGACCTTTAGTGAACCTTCAGGATTCCTGCAAAAGTTGTAGCCAGAC 274
DB 1556 GCTAAGTAAATTAAGTACTACTAGATATACTGTAACAAATTTGTTCAACCTAAAC 1615
QY 275 ATGTGATGATGCTTATCTGCATTATG 301
DB 1616 AATCTGTAATGCTTATTTTATG 1642

RESULT 11

US-08-985-335-4
; Sequence 4, Application US/08985335
; Patent No. 6080847
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi

APPLICANT: Corley, Neil C.
TITLE OF INVENTION: PROTEINS ASSOCIATED WITH CELL
PROLIFERATION
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,335
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0421 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2082 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LATRTUT02
CLONE: 1352286
US-08-985-335-4

Query Match 6.7%; Score 30; DB 3; Length 2082;
Best Local Similarity 50.0%; Pred. No. 2.7;
Matches 75; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 152 GTAAAGAGGAGACAAAGTAAATCTCTTTTAAAGTTTTCGCGTGGTGTCTAG 211
DB 1584 GGATAAGGTACAAATGTCGATTGTAATTAGTTTGTATTTCTATCTACTGCTGTAGA 1643
QY 212 TCACGACGATGACCTTTAGTGAACCTTTCGAGGATTTCTACGCAAGCTTTAGGCCAGA 271
DB 1644 AAACATATGTTTACATTTGATTAACTGTTGCTCTATTTATGCGAGGTAATCCAGCTAAA 1703
QY 272 GACATGTGATGATGCTTATCTGCAATTATG 301
DB 1704 GGAACCTTCTTTAATTATTAAGTATTATG 1733

RESULT 12
US-09-410-372-4
Sequence 4, Application US/09410372
Patent No. 6281334
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Yue, Henry
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purvi
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: PROTEINS ASSOCIATED WITH CELL
PROLIFERATION
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.

CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/410,372
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/985,335
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0421 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2082 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LATRTUT02
CLONE: 1352286
US-09-410-372-4

Query Match 6.7%; Score 30; DB 4; Length 2082;
Best Local Similarity 50.0%; Pred. No. 2.7;
Matches 75; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 152 GTAAAGAGGAGACAAAGTAAATCTCTTTTAAAGTTTTCGCGTGGTGTCTAG 211
DB 1584 GGATAAGGTACAAATGTCGATTGTAATTAGTTTGTATTTCTATCTACTGCTGTAGA 1643
QY 212 TCACGACGATGACCTTTAGTGAACCTTTCGAGGATTTCTACGCAAGCTTTAGGCCAGA 271
DB 1644 AAACATATGTTTACATTTGATTAACTGTTGCTCTATTTATGCGAGGTAATCCAGCTAAA 1703
QY 272 GACATGTGATGATGCTTATCTGCAATTATG 301
DB 1704 GGAACCTTCTTTAATTATTAAGTATTATG 1733

RESULT 13
US-08-694-869-3
Sequence 3, Application US/08694869
Patent No. 5994123
GENERAL INFORMATION:
APPLICANT: Olszewski, N.
APPLICANT: Tzafir, I.
APPLICANT: Somers, D. A.
APPLICANT: Lockhart, B.
APPLICANT: Torbert, K.
TITLE OF INVENTION: SUGARCANE BACILLIFORM VIRUS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
STREET: P.O. Box 2938
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible

```
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/694,869
; FILING DATE: 09-AUG-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 600.369US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-373-6900
; TELEFAX: 612-339-3061
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1207 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-08-694-869-3

Query Match 6.5%; Score 29.4; DB 2; Length 1207;
Best Local Similarity 70.9%; Pred. No. 3.3;
Matches 39; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 20 GAAGATGCTCTAGATCTGCTGGATATCAGTATGATGACTGAAGCGGAGTGCGG 74
DB 1112 GAAGATGCTCTAGATCTCCTCAAGACGTAAGCAATGACGATTGAGGAGGCATTGACG 1166

RESULT 14
US-09-349-546-3
; Sequence 3, Application US/09349546
; Patent No. 6093569
; GENERAL INFORMATION:
; APPLICANT: Olszewski, N.
; APPLICANT: Tzafir, I.
; APPLICANT: Somers, D.A.
; APPLICANT: Lockhart, B.
; APPLICANT: Torbert, K.
; TITLE OF INVENTION: Sugarcane bacilliform virus promoter
; FILE REFERENCE: 600.369US2
; CURRENT APPLICATION NUMBER: US/09/349,546
; CURRENT FILING DATE: 1999-07-08
; EARLIER APPLICATION NUMBER: US 08/694,869
; EARLIER FILING DATE: 1996-08-09
; EARLIER APPLICATION NUMBER: PCT/IB97/01338
; EARLIER FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1207
; TYPE: DNA
; ORGANISM: sugarcane bacilliform virus
US-09-349-546-3

Query Match 6.5%; Score 29.4; DB 3; Length 1207;
Best Local Similarity 70.9%; Pred. No. 3.3;
Matches 39; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 20 GAAGATGCTCTAGATCTGCTGGATATCAGTATGATGACTGAAGCGGAGTGCGG 74
DB 1112 GAAGATGCTCTAGATCTCCTCAAGACGTAAGCAATGACGATTGAGGAGGCATTGACG 1166

RESULT 15
US-08-694-869-4
; Sequence 4, Application US/08694869
```

```
; Patent No. 5994123
; GENERAL INFORMATION:
; APPLICANT: Olszewski, N.
; APPLICANT: Tzafir, I.
; APPLICANT: Somers, D.A.
; APPLICANT: Lockhart, B.
; APPLICANT: Torbert, K.
; TITLE OF INVENTION: SUGARCANE BACILLIFORM VIRUS
; TITLE OF INVENTION: PROMOTER
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/694,869
; FILING DATE: 09-AUG-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 600.369US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-373-6900
; TELEFAX: 612-339-3061
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1301 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-08-694-869-4

Query Match 6.5%; Score 29.4; DB 2; Length 1301;
Best Local Similarity 70.9%; Pred. No. 3.4;
Matches 39; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 20 GAAGATGCTCTAGATCTGCTGGATATCAGTATGATGACTGAAGCGGAGTGCGG 74
DB 1112 GAAGATGCTCTAGATCTCCTCAAGACGTAAGCAATGACGATTGAGGAGGCATTGACG 1166

Search completed: June 24, 2003, 05:36:39
Job time : 56 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 05:09:56 ; Search time 441 Seconds
(without alignments)
1500.702 Million cell updates/sec

Title: US-09-622-978-2

Perfect score: 451

Sequence: 1 agaacagaatatatttg.....acacactttgtcgagttca 451

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications: NA: *

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq: *
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq: *
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq: *
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq: *
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq: *
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq: *
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq: *
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq: *
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq: *
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq: *
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq: *
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq: *
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq: *
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	33.4	7.4	6071	9	US-10-239-676-77
C 2	32.4	7.2	197997	10	US-09-822-246-3
C 3	31.8	7.1	1180	10	US-09-881-752A-247
C 4	31.8	7.1	1234	10	US-09-956-004-81
5	31.6	7.0	214	9	US-09-997-279-94
6	31.6	7.0	214	10	US-09-813-358-94
C 7	31.6	7.0	460	10	US-09-864-761-10041
C 8	31.4	7.0	1456	9	US-09-938-842A-4457
C 9	31	6.9	17056	10	US-09-993-238-3
C 10	30.6	6.8	378361	9	US-09-901-136-3
C 11	30.4	6.7	1176	9	US-09-992-598-413
C 12	30.4	6.7	1176	9	US-09-989-293A-413
C 13	30.4	6.7	1176	9	US-10-063-547-87
C 14	30.4	6.7	1176	9	US-09-989-735-413
C 15	30.4	6.7	1176	9	US-09-990-444-413
C 16	30.4	6.7	1176	9	US-09-989-730-413
C 17	30.4	6.7	1176	9	US-09-990-436-413
C 18	30.4	6.7	1176	9	US-09-991-181-413
C 19	30.4	6.7	1176	9	US-09-993-687-413

ALIGNMENTS

RESULT 1

US-10-239-676-77/C

; Sequence 77, Application US/10239676

; Publication No. US20030082609A1

; GENERAL INFORMATION:

; APPLICANT: OLEK, Alexander

; APPLICANT: PIERBROCK, Christian

; APPLICANT: BERLIN, Kurt

; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation

; FILE REFERENCE: 5013.1003

; CURRENT APPLICATION NUMBER: US/10/239,676

; CURRENT FILING DATE: 2002-09-24

; PRIOR APPLICATION NUMBER: PCT/EP01/03968

; DE 10019058.8

; DE 10019173.8

; DE 10032529.7

; DE 10043826.1

; PRIOR FILING DATE: 2001-04-06

; 2000-04-06

; 2000-04-07

; 2000-06-30

; 2000-09-01

; NUMBER OF SEQ ID NOS: 228

; SEQ ID NO 77

; LENGTH: 6071

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

; NAME/KEY: unsure

; LOCATION: (3292, 3454)

US-10-239-676-77

Query Match

Best Local Similarity 7.4%; Score 33.4; DB 9; Length 6071;

Matches 97; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 215 GCACGATGACCTTGTAGTACCTTTCAGGATCTTACGCAAGTTGTATGGCCAGAC 274

DB 4118 GCACCAACGCAATAACACAAAATAATAATTAATTCTTCAAAACGCTCTCAATACGAC 4059


```
QY 129 GAGCTGATACCACTCAGCTTTATGTAAGAGAGAGACAAAGTATAATGCTCTCTTTATTTA 188
    || || || || || || || || || || || || || || || || || || || || || ||
Db 356 GACTTTATGAACCCACTNATGATATGATATGATATATTTCTCTCTTCATTTATA 297
    || || || || || || || || || || || || || || || || || || || || || ||
QY 189 AGTTTGTGGTGTGCTGTCTAGTCAC 215
    || || || || || || || || || || || || || || || || || || || || || ||
Db 296 ATTTTCTTATAACATTTCTTTATCTC 270
    || || || || || || || || || || || || || || || || || || || || || ||

RESULT 8
US-09-938-842A-4457
; Sequence 4457, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4457
; LENGTH: 1456
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4457

Query Match 7.0%; Score 31.4; DB 9; Length 1456;
Best Local Similarity 50.2%; Pred. No. 7.9;
Matches 103; Conservative 0; Mismatches 101; Indels 1; Gaps 1;

QY 80 CTACACGCTGTGATACCAACCGGTGTGAAGCTGATAGATCGGAGTGAGTGATAC 139
    || || || || || || || || || || || || || || || || || || || || || ||
Db 911 CCACATGTTTGTGTTATGAAATGATTAATCAATCAACCGCCTTCATCTCAAGTTAA 970
    || || || || || || || || || || || || || || || || || || || || || ||
QY 140 CACTCAGTTATGTAAGAGAGAGACAAAGTATAATGCTCTTTATTTAAGTTTGTGCT 199
    || || || || || || || || || || || || || || || || || || || || || ||
Db 971 AAATCAATCTTAATTAGTGAAGACAAAGAAAGACATATTTTAAAGTGAGTTCTC 1030
    || || || || || || || || || || || || || || || || || || || || || ||
QY 200 GTCGTTGTCTAGTCACGACGATGACCTTTAGTGAAGTTCGAGGATCTTAC-GCAAG 258
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1031 TCTCTGTACACTCACAAAGCTTTATCTTTCTGGCTACTGCAAGCTCATCTGAAAG 1090
    || || || || || || || || || || || || || || || || || || || || || ||

QY 259 TTGTTAGGCCAGACATGTGATGA 283
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1091 AGCTTAACCAAGAGATCTGTATA 1115
    || || || || || || || || || || || || || || || || || || || || || ||

RESULT 9
US-09-933-238-3/c
; Sequence 3, Application US/09893238
; Patent No. US20020150973A1
; GENERAL INFORMATION:
; APPLICANT: Moore, K.
; APPLICANT: Nagle, D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
; FILE REFERENCE: 7853-237
; CURRENT APPLICATION NUMBER: US/09/893,238
; CURRENT FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: 09/245,041
; PRIOR FILING DATE: 1999-02-05
; PRIOR APPLICATION NUMBER: 60/093,630
; PRIOR FILING DATE: 1998-07-21
```

```
; PRIOR APPLICATION NUMBER: 60/104,978
; PRIOR FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 17056
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-893-238-3

Query Match 6.9%; Score 31; DB 10; Length 17056;
Best Local Similarity 62.0%; Pred. No. 42;
Matches 49; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 136 ATACCACCTCATTATGTAAGAGAGAGACAAAGTATAATGCTCTTTATTTAAGTTTGT 195
    || || || || || || || || || || || || || || || || || || || || || ||
Db 12090 ATGCCAAGCACATTTCTCTATCTAGAACTAAAGTGATAATGCTATCAATTTAAATTTG 12031
    || || || || || || || || || || || || || || || || || || || || || ||

QY 196 CGGTGCTGTGTCTAGTCA 214
    || || || || || || || || || || || || || || || || || || || || || ||
Db 12030 TGGTGGGTTGTAGTAA 12012
    || || || || || || || || || || || || || || || || || || || || || ||

RESULT 10
US-09-901-136-3/c
; Sequence 3, Application US/09901136
; Publication No. US20030039968A1
; GENERAL INFORMATION:
; APPLICANT: HU, Song et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001273
; CURRENT APPLICATION NUMBER: US/09/901,136
; CURRENT FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 378361
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(378361)
; OTHER INFORMATION: n = A,T,C or G
US-09-901-136-3

Query Match 6.8%; Score 30.6; DB 9; Length 378361;
Best Local Similarity 51.9%; Pred. No. 3.1e+02;
Matches 69; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 184 TTTTAAAGTTTGTGGTGCTGCTGTTAGTCAAGCAGGATGACCTTTAGTGAACCTTGCAG 243
    || || || || || || || || || || || || || || || || || || || || || ||
Db 267381 TTTCATCACAGTTGTGCTGATTTGTGATACATCATCATCTGTTCTGAAATGAGGATACAA 267322
    || || || || || || || || || || || || || || || || || || || || || ||

QY 244 GATTCTTACGCAAGTTGTTAGGCCAGACAGATGTGATGCTTATCTGTCATTATTTG 303
    || || || || || || || || || || || || || || || || || || || || || ||
Db 267321 TTTTGTACTCAAGTTATATTTCCAAAGCTATATTTTAACTACACCTTCAGTCTTGT 267262
    || || || || || || || || || || || || || || || || || || || || || ||

QY 304 GGATGCCACCTAA 316
    || || || || || || || || || || || || || || || || || || || || || ||
Db 267261 TATTGCATGGTAA 267249
    || || || || || || || || || || || || || || || || || || || || || ||

RESULT 11
US-09-992-598-413
; Sequence 413, Application US/09992598
; Patent No. US20020160384A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
```

```

; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pad, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2730PIC20
; CURRENT APPLICATION NUMBER: US/09/992,598
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/065770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088555
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444

```

;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090445
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090472
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090535
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090540
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090542
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090676
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090678
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090690
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090694
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090695
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090696
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090862
;; PRIOR FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: 60/090863
;; PRIOR FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: 60/091360
;; PRIOR FILING DATE: 1998-07-01
;; PRIOR APPLICATION NUMBER: 60/091478
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091544
;; PRIOR FILING DATE: 1998-07-01
;; PRIOR APPLICATION NUMBER: 60/091519
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091626
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091633
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091978
;; PRIOR FILING DATE: 1998-07-07
;; PRIOR APPLICATION NUMBER: 60/091982
;; PRIOR FILING DATE: 1998-07-07
;; PRIOR APPLICATION NUMBER: 60/092182
;; PRIOR FILING DATE: 1998-07-09

Query Match 6.7%; Score 30.4; DB 9; Length 1176;
Best Local Similarity 53.3%; Pred. No. 15;
Matches 64; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
Qy 43 TATCAGTATGATGACTGAGCGGAGTGGCGGACCCCTACACGTGTGTATACCAACCG 102
Db 879 TAACTGAGCATCACTGCTGTTGGAGGAGGATCTTCCAGAGGCCAGTCCCCAGCA 938
Qy 103 GTGTGAAGACTGATAAGATGCGGAGTGAGCTGGATACCACTCACCTTTATGTAAGAGGAG 162
Db 939 GTGTGGAGATTTCTTCGTTTGTGTTGATGGAGTGATGGAATCATGTGTTACAGCAG 998

RESULT 12

US-09-989-293A-413
; Sequence 413, Application US/09989293A
; Patent No. US2002017164A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman

;; APPLICANT: Gerber, Hanspeter
;; APPLICANT: Gerritsen, Mary E.
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Grimaldi, J. Christopher
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Kljavin, Ivar J.
;; APPLICANT: Napier, Mary A.
;; APPLICANT: Pan, James
;; APPLICANT: Paoni, Nicholas F.
;; APPLICANT: Roy, Margaret Ann
;; APPLICANT: Stewart, Timothy A.
;; APPLICANT: Tumas, Daniel
;; APPLICANT: Watanabe, Colin K.
;; APPLICANT: Williams, P. Mickey
;; APPLICANT: Wood, William I.
;; APPLICANT: Zhang, Zemin
;; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;; FILE REFERENCE: P2730PIC66
;; CURRENT APPLICATION NUMBER: US/09/989,293A
;; CURRENT FILING DATE: 2001-11-20
;; PRIOR APPLICATION NUMBER: 60/049787
;; PRIOR FILING DATE: 1997-06-16
;; PRIOR APPLICATION NUMBER: 60/062250
;; PRIOR FILING DATE: 1997-10-17
;; PRIOR APPLICATION NUMBER: 60/065186
;; PRIOR FILING DATE: 1997-11-12
;; PRIOR APPLICATION NUMBER: 60/065311
;; PRIOR FILING DATE: 1997-11-13
;; PRIOR APPLICATION NUMBER: 60/066770
;; PRIOR FILING DATE: 1997-11-24
;; PRIOR APPLICATION NUMBER: 60/075945
;; PRIOR FILING DATE: 1998-02-25
;; PRIOR APPLICATION NUMBER: 60/078910
;; PRIOR FILING DATE: 1998-03-20
;; PRIOR APPLICATION NUMBER: 60/083322
;; PRIOR FILING DATE: 1998-04-28
;; PRIOR APPLICATION NUMBER: 60/084500
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/087106
;; PRIOR FILING DATE: 1998-05-28
;; PRIOR APPLICATION NUMBER: 60/087607
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/087609
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/087759
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/087827
;; PRIOR FILING DATE: 1998-06-03
;; PRIOR APPLICATION NUMBER: 60/088021
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088025
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088026
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088028
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088029
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088030
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088033
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088326
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088167
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088202
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088212
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088217

1	PRIOR FILING DATE: 1998-06-05
2	PRIOR APPLICATION NUMBER: 60/088655
3	PRIOR FILING DATE: 1998-06-09
4	PRIOR APPLICATION NUMBER: 60/088734
5	PRIOR FILING DATE: 1998-06-10
6	PRIOR APPLICATION NUMBER: 60/088738
7	PRIOR FILING DATE: 1998-06-10
8	PRIOR APPLICATION NUMBER: 60/088742
9	PRIOR FILING DATE: 1998-06-10
10	PRIOR APPLICATION NUMBER: 60/088810
11	PRIOR FILING DATE: 1998-06-10
12	PRIOR APPLICATION NUMBER: 60/088824
13	PRIOR FILING DATE: 1998-06-10
14	PRIOR APPLICATION NUMBER: 60/088826
15	PRIOR FILING DATE: 1998-06-10
16	PRIOR APPLICATION NUMBER: 60/088858
17	PRIOR FILING DATE: 1998-06-11
18	PRIOR APPLICATION NUMBER: 60/088861
19	PRIOR FILING DATE: 1998-06-11
20	PRIOR APPLICATION NUMBER: 60/088876
21	PRIOR FILING DATE: 1998-06-11
22	PRIOR APPLICATION NUMBER: 60/089105
23	PRIOR FILING DATE: 1998-06-12
24	PRIOR APPLICATION NUMBER: 60/089440
25	PRIOR FILING DATE: 1998-06-16
26	PRIOR APPLICATION NUMBER: 60/089512
27	PRIOR FILING DATE: 1998-06-16
28	PRIOR APPLICATION NUMBER: 60/089514
29	PRIOR FILING DATE: 1998-06-16
30	PRIOR APPLICATION NUMBER: 60/089532
31	PRIOR FILING DATE: 1998-06-17
32	PRIOR APPLICATION NUMBER: 60/089538
33	PRIOR FILING DATE: 1998-06-17
34	PRIOR APPLICATION NUMBER: 60/089598
35	PRIOR FILING DATE: 1998-06-17
36	PRIOR APPLICATION NUMBER: 60/089599
37	PRIOR FILING DATE: 1998-06-17
38	PRIOR APPLICATION NUMBER: 60/089600
39	PRIOR FILING DATE: 1998-06-17
40	PRIOR APPLICATION NUMBER: 60/089653
41	PRIOR FILING DATE: 1998-06-17
42	PRIOR APPLICATION NUMBER: 60/089801
43	PRIOR FILING DATE: 1998-06-18
44	PRIOR APPLICATION NUMBER: 60/089907
45	PRIOR FILING DATE: 1998-06-18
46	PRIOR APPLICATION NUMBER: 60/089908
47	PRIOR FILING DATE: 1998-06-18
48	PRIOR APPLICATION NUMBER: 60/089947
49	PRIOR FILING DATE: 1998-06-19
50	PRIOR APPLICATION NUMBER: 60/089948
51	PRIOR FILING DATE: 1998-06-19
52	PRIOR APPLICATION NUMBER: 60/089952
53	PRIOR FILING DATE: 1998-06-19
54	PRIOR APPLICATION NUMBER: 60/090246
55	PRIOR FILING DATE: 1998-06-22
56	PRIOR APPLICATION NUMBER: 60/090252
57	PRIOR FILING DATE: 1998-06-22
58	PRIOR APPLICATION NUMBER: 60/090254
59	PRIOR FILING DATE: 1998-06-22
60	PRIOR APPLICATION NUMBER: 60/090349
61	PRIOR FILING DATE: 1998-06-23
62	PRIOR APPLICATION NUMBER: 60/090355
63	PRIOR FILING DATE: 1998-06-23
64	PRIOR APPLICATION NUMBER: 60/090429
65	PRIOR FILING DATE: 1998-06-24
66	PRIOR APPLICATION NUMBER: 60/090431
67	PRIOR FILING DATE: 1998-06-24
68	PRIOR APPLICATION NUMBER: 60/090435
69	PRIOR FILING DATE: 1998-06-24
70	PRIOR APPLICATION NUMBER: 60/090444
71	PRIOR FILING DATE: 1998-06-24
72	PRIOR APPLICATION NUMBER: 60/090445
73	PRIOR FILING DATE: 1998-06-24

3	PRIOR APPLICATION NUMBER: 60/090472
3	PRIOR FILING DATE: 1998-06-24
3	PRIOR APPLICATION NUMBER: 60/090535
3	PRIOR FILING DATE: 1998-06-24
3	PRIOR APPLICATION NUMBER: 60/090540
3	PRIOR FILING DATE: 1998-06-24
3	PRIOR APPLICATION NUMBER: 60/090542
3	PRIOR FILING DATE: 1998-06-24
3	PRIOR APPLICATION NUMBER: 60/090557
3	PRIOR FILING DATE: 1998-06-24
3	PRIOR APPLICATION NUMBER: 60/090676
3	PRIOR FILING DATE: 1998-06-25
3	PRIOR APPLICATION NUMBER: 60/090678
3	PRIOR FILING DATE: 1998-06-25
3	PRIOR APPLICATION NUMBER: 60/090690
3	PRIOR FILING DATE: 1998-06-25
3	PRIOR APPLICATION NUMBER: 60/090694
3	PRIOR FILING DATE: 1998-06-25
3	PRIOR APPLICATION NUMBER: 60/090695
3	PRIOR FILING DATE: 1998-06-25
3	PRIOR APPLICATION NUMBER: 60/090696
3	PRIOR FILING DATE: 1998-06-25
3	PRIOR APPLICATION NUMBER: 60/090862
3	PRIOR FILING DATE: 1998-06-26
3	PRIOR APPLICATION NUMBER: 60/090863
3	PRIOR FILING DATE: 1998-06-26
3	PRIOR APPLICATION NUMBER: 60/091360
3	PRIOR FILING DATE: 1998-07-01
3	PRIOR APPLICATION NUMBER: 60/091478
3	PRIOR FILING DATE: 1998-07-02
3	PRIOR APPLICATION NUMBER: 60/091544
3	PRIOR FILING DATE: 1998-07-01
3	PRIOR APPLICATION NUMBER: 60/091519
3	PRIOR FILING DATE: 1998-07-02
3	PRIOR APPLICATION NUMBER: 60/091626
3	PRIOR FILING DATE: 1998-07-02
3	PRIOR APPLICATION NUMBER: 60/091633
3	PRIOR FILING DATE: 1998-07-02
3	PRIOR APPLICATION NUMBER: 60/091978
3	PRIOR FILING DATE: 1998-07-07
3	PRIOR APPLICATION NUMBER: 60/091982
3	PRIOR FILING DATE: 1998-07-07
3	PRIOR APPLICATION NUMBER: 60/092182
3	PRIOR FILING DATE: 1998-07-09

Query Match 6.7%; Score 30.4; DB 9; Length 1176;

Best Local Similarity	53.38;	Pred. No. 15;		
Matches	64;	Conservative	0;	Mismatches
			56;	Indels
				0;
				Gaps
				0;

43 TATCAGTAATGATGACTGAAGCGGAGTGCGGGACCCCTACCACGTGTGATACCAACCG 102 QY

Db 879 TAACACTGAGCATCACTGCATTGGTGGAGGAGGATACTTCCAGAGGCCAGTCCCCAGCA 938

103 GTGTGAAGACTGATAAGATGCGGAGTGAGCTGGATACCACTCATTATGTAAAGAGGAG 162

939 CTCTGGAGATTTTTCTCGGTTTGATGGAGTCGATATGGAACTCATCTTCTCTTACCCAG 988

RESULT 13

```

RESULTS 13
US-10-063-547-87
; Sequence 87, Application US/10063547
; Publication No. US20020192638A1
; GENERAL INFORMATION:
; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvarcoff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey J.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grunaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRA

```

[illegible]

PRIOR APPLICATION NUMBER: 60/089651
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090255
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090535
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090540
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090542
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090676
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090678
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090690
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090694
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090695
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090696
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090862
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/090863
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/091360
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091478
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091544
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091519
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091626
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091633
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091978

Search completed: June 24, 2003, 06:45:32
Job time : 443 secs